

THE ~~Q~~UARTERLY JOURNAL OF
MECHANICS AND
APPLIED
MATHEMATICS

VOLUME X PART 3

AUGUST 1957

UNIVERSITY
OF MICHIGAN

AUG 12 1957

ENGINEERING
LIBRARY

OXFORD
AT THE CLARENDON PRESS
1957

Price 18s. net

PRINTED IN GREAT BRITAIN BY CHARLES BATEY AT THE UNIVERSITY PRESS, OXFORD

THE QUARTERLY JOURNAL OF MECHANICS AND APPLIED MATHEMATICS

Editorial Board

D. G. CHRISTOPHERSON L. HOWARTH
G. I. TAYLOR G. TEMPLE

together with

A. C. AITKEN	M. J. LIGHTHILL
S. CHAPMAN	G. C. McVITTIE
A. R. COLLAR	N. F. MOTT
T. G. COWLING	W. G. PENNEY
C. G. DARWIN	A. G. PUGSLEY
W. J. DUNCAN	L. ROSENHEAD
S. GOLDSTEIN	R. V. SOUTHWELL
A. E. GREEN	O. G. SUTTON
A. A. HALL	ALEXANDER THOM
D. R. HARTREE	A. H. WILSON
WILLIS JACKSON	J. R. WOMERSLEY
H. JEFFREYS	

Executive Editors

V. C. A. FERRARO D. M. A. LEGGETT

THE QUARTERLY JOURNAL OF MECHANICS AND APPLIED MATHEMATICS is published at 18s. net for a single number with an annual subscription (for four numbers) of 60s. post free.

NOTICE TO CONTRIBUTORS

1. *Communication.* Papers should be communicated to Dr. D. M. A. Leggett, Department of Mathematics, King's College, Strand, London, W.C. 2.

If possible, to expedite publication, papers should be submitted in duplicate.

2. *Presentation.* Papers should be typewritten (double spacing) and should be preceded by a summary not exceeding 300 words in length. References to literature should be given in standard order, *author, title of journal, volume number, date, page*. These should be placed at the end of the paper and arranged according to the order of reference in the paper.

3. *Diagrams.* The number of diagrams should be kept to the minimum consistent with clarity. The lines of the figures should be drawn in ink either on draughtsman's paper or on good quality white paper. Each individual line in the figure should bear reducing to one-half of the size of the original, and great care should be exercised to see that the lines are regular in thickness, especially where they meet. Lettering of the figure should be in pencil and should be sufficient to define clearly the lines and curves in it. The writing of formulae or of explanations on the diagram itself should be avoided. All explanations of symbols, etc., should be given in underline. Contributors should indicate on their manuscripts where figures should be inserted.

4. *Tables.* Tables should preferably be arranged so that they can be printed with the columns parallel to the longer edge of the page.

5. *Notation.* All single letters used to denote vectors in the manuscript should be marked by underlining with a wavy line. Scalar and vector products should be denoted by $\underline{g} \cdot \underline{h}$ and $\underline{g} \wedge \underline{h}$ respectively. Real and imaginary parts of complex quantities should be denoted by *re* and *im* respectively.

6. *Offprints.* Authors of papers will be entitled to 25 free offprints. This number is available for sharing between authors of joint papers.

7. All correspondence other than that dealing with contributions should be addressed to the publishers:

OXFORD UNIVERSITY PRESS
AMEN HOUSE, LONDON, E.C. 4

S

is
or

rt-

ed
be
se
of

th
er
ng
he
ld
he
all
ld

he

be
ed
ld

is

ed

E

By
dynam
over th
special
blast lo

1. In

ELAST

the su

Sezaw

(4), G.

throw

surface

partic

solutio

includ

harmo

A fe

dimen

from t

the sp

repres

sive se

assum

Whi

waves

tant s

loading

ables.

is intri

† Pro

‡ Ref

§ The

reference

[Quart.

5002.36

ELASTO-DYNAMIC PROBLEM CONCERNING THE SPHERICAL CAVITY

By A. CEMAL ERINGEN†
(Purdue University, Lafayette, Indiana)

[Received 15 June 1956]

SUMMARY

By using the Fourier transform technique the solution is given for the elastodynamic problem resulting from the application of arbitrary dynamical tractions over the surface of a spherical cavity in an infinite, isotropic, elastic solid. Various special cases are studied. Computations are carried out and plotted in the case of blast loading.

1. Introduction

ELASTIC wave propagations originating from the application of forces on the surface of a spherical cavity have been studied by K. Sezawa (1), K. Sezawa and K. Kanai (2), W. Inouye (3), H. Kawasumi and R. Yosiyama (4), G. Nishimura (5), and Vaněk (6). The basic aim of these studies was to throw light on the origin of earthquake phenomena. Consequently, the surface traction distribution on the surface of the cavity was limited to some particular types. Moreover, attention was directed to the behaviour of the solution at great distances from the cavity. Basically the cases studied include the azimuthal distribution of the type $P_0(\cos \theta)$, $P_2(\cos \theta)$ with harmonic or exponential type of loading in time.‡

A few recent papers (7-10)§ have been devoted to the study of the one-dimensional problem, namely, the propagation of elastic waves originating from the application of uniform time-dependent pressure on the surface of the spherical cavity. Of these, the papers by Allen and Goldsmith (7, 8) represent a valuable contribution to this problem on account of the extensive set of curves for the displacement, velocity, and stresses. These papers assume an exponentially decaying pressure in time.

While the above studies have important application to the problem of waves generated by earthquakes and explosion, they exclude some important special cases of loading, such as blast, concentrated impact, ring loading, and the general case of arbitrary loading in space and time variables. Also the behaviour of the solution in the neighbourhood of the cavity is intricate and requires further attention.

† Professor of Engineering Sciences.

‡ References (1-5) were discovered upon the completion of the analytical work, Dec. 1955.

§ The author is indebted to a referee for calling his attention to these references and to reference (6).

The present paper deals with the last-mentioned class of problems which appear to be basic theoretically and are important from the point of view of applications in engineering. By using the Fourier transform we give the general solution of the elasto-dynamic problem for a spherical cavity under the action of arbitrary dynamical surface tractions. The solution given in all of the foregoing papers may, of course, be obtained as special cases of the present solution.

Various particular cases of practical importance are studied. Among these, the uniformly applied normal dynamic load leads to a closed form solution. Curiously enough, this is not the case for the analogous problem in two dimensions, namely, the circular hole in an infinite plate.[†] Computations are carried out for the case of blast and discussed in the body of the paper.

2. Formulation of the problem

The equations of small motion of a homogeneous, isotropic, elastic body under no body forces are

$$(\lambda + \mu)\nabla(\nabla \cdot \mathbf{u}) + \mu\nabla^2 \mathbf{u} = \rho \mathbf{u}_{,tt}. \quad (1)$$

Here \mathbf{u} is the displacement vector; λ , μ are the Lamé constants, ρ is the mass density, ∇ is the gradient operator, $\nabla \cdot$ and ∇^2 represent the divergence and Laplacian; t is the time, and the comma indicates differentiation, i.e. $\mathbf{u}_{,tt} = \partial^2 \mathbf{u} / \partial t^2$. In spherical coordinates (r, θ, ϕ) (Fig. 1), \mathbf{u} has the components u, v, w . Components of stress tensor $\sigma_r, \dots, \tau_{\theta\phi}$ are related to displacement components u, v, w by

$$\begin{aligned} \sigma_r &= \lambda\Delta + 2\mu u_{,r}, \\ \sigma_\theta &= \lambda\Delta + 2\mu r^{-1}(u + v_{,\theta}), \\ \sigma_\phi &= \lambda\Delta + 2\mu r^{-1}(u + v \cot \theta + w_{,\phi} \operatorname{cosec} \theta), \\ \tau_{r\theta} &= \mu(v_{,r} - r^{-1}v + r^{-1}u_{,\theta}), \\ \tau_{r\phi} &= \mu[w_{,r} - r^{-1}w + (r \sin \theta)^{-1}u_{,\phi}], \\ \tau_{\theta\phi} &= \mu r^{-1}(w_{,\theta} - w \cot \theta + v_{,\phi} \operatorname{cosec} \theta), \end{aligned} \quad (2)$$

$$\text{where} \quad \Delta \equiv \nabla \cdot \mathbf{u} = u_{,r} + r^{-1}[2u + v_{,\theta} + v \cot \theta + (\sin \theta)^{-1}w_{,\phi}] \quad (3)$$

is the dilatation.

The problem is now reduced to finding the solution of (1) which vanishes at $r = \infty$ and at the surface of the spherical cavity, $r = a$, satisfies the boundary conditions

$$\begin{aligned} \sigma_r(a, \theta, \phi, t) &= R(\theta, \phi, t), \\ \tau_{r\theta}(a, \theta, \phi, t) &= \Theta(\theta, \phi, t), \\ \tau_{r\phi}(a, \theta, \phi, t) &= \Phi(\theta, \phi, t). \end{aligned} \quad (4)$$

[†] These results are now being submitted for publication.

It is further required that the waves generated at $r = a$ will be of outgoing type.

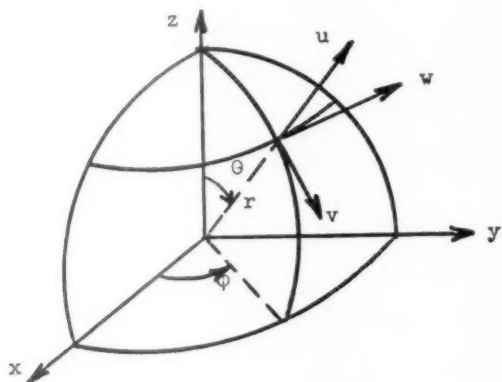


FIG. 1. Spherical coordinates

3. The solution

We use the Fourier transform technique to solve the problem stated. The Fourier transform $\bar{f}(p)$ of $f(t)$, and $f(t)$, are related to each other by the equations

$$\bar{f}(p) = \int_{-\infty}^{\infty} f(t)e^{ipt} dt, \quad f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \bar{f}(p)e^{-ipt} dp. \quad (5)$$

Multiplying both sides of (1) by e^{ipt} and integrating the result from $-\infty$ to ∞ we obtain

$$(\nabla^2 + h_2^2)\bar{\mathbf{u}} = (1 - h_2^2/h_1^2)\nabla(\nabla \cdot \bar{\mathbf{u}}), \quad (6)$$

$$\text{where } h_1^2 = \rho p^2/(\lambda + 2\mu) = \alpha^2 p^2, \quad h_2^2 = \rho p^2/\mu = \beta^2 p^2. \quad (7)$$

Taking the divergence of both sides of (6) we obtain

$$(\nabla^2 + h_1^2)\nabla \cdot \bar{\mathbf{u}} = 0. \quad (8)$$

Assuming that $\nabla \cdot \bar{\mathbf{u}}$ is determined from (8) a particular solution $\bar{\mathbf{u}}_1$ of (6) is

$$\bar{\mathbf{u}}_1 = h_1^{-2}\nabla(\nabla \cdot \bar{\mathbf{u}}). \quad (9)$$

A more complete solution of (6) is obtained by adding to $\bar{\mathbf{u}}_1$ the complementary solution $\bar{\mathbf{u}}_2$ of equations

$$(\nabla^2 + h_2^2)\bar{\mathbf{u}}_2 = 0, \quad \nabla \cdot \bar{\mathbf{u}}_2 = 0. \quad (10)$$

The desired solution must satisfy the boundary conditions at $r = a$ and

$r = \infty$ and must have the character of diverging waves. The latter condition is guaranteed by the Sommerfeld radiation condition, i.e.

$$\lim_{r \rightarrow \infty} r(\bar{\mathbf{u}}_{k,r} - i h_k \bar{\mathbf{u}}_k) = 0 \quad (k = 1, 2). \quad (11)$$

A solution of (8) may be obtained in terms of the spherical surface harmonics S_n of positive integral degree n in the form

$$\nabla \cdot \bar{\mathbf{u}}_1 = r^n R_n(r) S_n, \quad (12)$$

where $R_n(r)$ is a function of r only which must satisfy the equation

$$R_{n,rr} + 2(n+1)r^{-1}R_{n,r} + h_1^2 R_n = 0. \quad (13)$$

The solution of this equation is given by

$$R_n = r^{-(n+\frac{1}{2})} [A H_{n+\frac{1}{2}}^{(1)}(h_1 r) + B H_{n+\frac{1}{2}}^{(2)}(h_1 r)], \quad (14)$$

where A and B are arbitrary constants and $H_{n+\frac{1}{2}}^{(1)}$ and $H_{n+\frac{1}{2}}^{(2)}$ are the half order Hankel functions. Of these functions only $H_{n+\frac{1}{2}}^{(1)}$ satisfies the Sommerfeld radiation condition, i.e. the asymptotic form of $H_{n+\frac{1}{2}}^{(1)}$ corresponds to outgoing waves since

$$H_{n+\frac{1}{2}}^{(1)}(h_k \rho) e^{-ipt} \sim (2/\pi h_k \rho)^{\frac{1}{2}} \exp[-i(n+1)\pi/2] \exp[i(h_k \rho - pt)].$$

Hence we set $B = 0$.

Also, absorbing A into S_n we write

$$\nabla \cdot \bar{\mathbf{u}}_1 = r^{-\frac{1}{2}} H_{n+\frac{1}{2}}^{(1)}(h_1 r) S_n. \quad (15)$$

The expression of $\bar{\mathbf{u}}_1$ now follows from (9):

$$\bar{\mathbf{u}}_1 = -h_1^{-2} \nabla(r^{-\frac{1}{2}} H_{n+\frac{1}{2}}^{(1)}(h_1 r) S_n). \quad (16)$$

The components $\bar{u}_1, \bar{v}_1, \bar{w}_1$ of $\bar{\mathbf{u}}_1$ are therefore given by:

$$\begin{aligned} \bar{u}_1 &= -h_1^{-2} [r^{-\frac{1}{2}} H_{n+\frac{1}{2}}^{(1)}(h_1 r) S_n]_{,r}, \\ \bar{v}_1 &= -h_1^{-2} [r^{-\frac{1}{2}} H_{n+\frac{1}{2}}^{(1)}(h_1 r) S_n]_{,\theta}, \\ \bar{w}_1 &= -h_1^{-2} [r^{-\frac{1}{2}} H_{n+\frac{1}{2}}^{(1)}(h_1 r) S_n]_{,\phi} \csc \theta. \end{aligned} \quad (17)$$

We now proceed to determine a solution $\bar{\mathbf{u}}_2$ of (10). Let

$$\bar{\mathbf{u}}_2 = R_n \mathbf{Z}_n, \quad (18)$$

where \mathbf{Z}_n is a vector point function having solid spherical harmonics as rectangular components, and R_n has the same form as (14) with $B = 0$ and h_2 replacing h_1 . The function

$$\mathbf{Z}_n = (\nabla W_n) \times \mathbf{r} \quad (19)$$

satisfies both equations (10). Here W_n is a spherical surface harmonic of degree n . Hence a solution of (10) is of the form

$$\bar{\mathbf{u}}_2 = r^{-n-\frac{1}{2}} H_{n+\frac{1}{2}}^{(1)}(h_2 r) (\nabla W_n) \times \mathbf{r}, \quad (20)$$

the components $\bar{u}_2, \bar{v}_2, \bar{w}_2$ of which are:

$$\begin{aligned}\bar{u}_2 &= 0, \\ \bar{v}_2 &= [r^{-\frac{1}{2}} H_{n+\frac{1}{2}}^{(1)}(h_2 r) W_n]_{,\phi} \operatorname{cosec} \theta, \\ \bar{w}_2 &= -[r^{-\frac{1}{2}} H_{n+\frac{1}{2}}^{(1)}(h_2 r) W_n]_{,\theta}.\end{aligned}\quad (21)$$

A second solution of (10) may be obtained by observing the fact that

$$\bar{\mathbf{u}}'_2 = \nabla \times \bar{\mathbf{u}}_2 = \nabla \times (R_n \nabla W_n \times \mathbf{r}) \quad (22)$$

satisfies (10). Hence in this equation if we replace W_n by another spherical surface harmonic Y_n of integral degree n we obtain another solution of (10):

$$\begin{aligned}\bar{u}'_2 &= r^{-\frac{1}{2}} H_{n+\frac{1}{2}}^{(1)}(h_2 r) n(n+1) Y_n, \\ \bar{v}'_2 &= r^{-1} [r^{\frac{1}{2}} H_{n+\frac{1}{2}}^{(1)}(h_2 r) Y_n]_{,r\theta}, \\ \bar{w}'_2 &= r^{-1} [r^{\frac{1}{2}} H_{n+\frac{1}{2}}^{(1)}(h_2 r) Y_n]_{,r\phi} \operatorname{cosec} \theta.\end{aligned}\quad (23)$$

The diverging wave type of solution of (6) which vanishes at $r = \infty$ is then given by the sum $\bar{\mathbf{u}} = \bar{\mathbf{u}}_1 + \bar{\mathbf{u}}_2 + \bar{\mathbf{u}}'_2$, whose components in terms of arbitrary spherical surface harmonics S_n, W_n, Y_n of integral degree n are

$$\begin{aligned}\bar{u} &= -h_1^{-2} [r^{-\frac{1}{2}} H_{n+\frac{1}{2}}^{(1)}(h_1 r) S_n]_{,r} + r^{-\frac{1}{2}} H_{n+\frac{1}{2}}^{(1)}(h_2 r) n(n+1) Y_n, \\ \bar{v} &= \{-h_1^{-2} r^{-\frac{3}{2}} H_{n+\frac{1}{2}}^{(1)}(h_1 r) S_n + r^{-1} [r^{\frac{1}{2}} H_{n+\frac{1}{2}}^{(1)}(h_2 r) Y_n]_{,r}\}_{,\theta} + \\ &\quad + [r^{-\frac{1}{2}} H_{n+\frac{1}{2}}^{(1)}(h_2 r) W_n]_{,\phi} \operatorname{cosec} \theta, \\ \bar{w} &= \{-h_1^{-2} r^{-\frac{3}{2}} H_{n+\frac{1}{2}}^{(1)}(h_1 r) S_n + r^{-1} [r^{\frac{1}{2}} H_{n+\frac{1}{2}}^{(1)}(h_2 r) Y_n]_{,r}\}_{,\phi} \operatorname{cosec} \theta - \\ &\quad - [r^{-\frac{1}{2}} H_{n+\frac{1}{2}}^{(1)}(h_2 r) W_n]_{,\theta}.\end{aligned}\quad (24)$$

Equation (24) satisfies (6) for every n . Hence any linear combination will also be a solution of (6). By summing over n from 0 to ∞ in each equation of (24) we can express the motion resulting from any dynamic traction at the surface of the cavity. This requires the computation of stress components which are obtained by substituting (24) into the Fourier transforms of (2). Hence

$$\begin{aligned}\bar{\sigma}_r &= \mu [A_n(pr) S_n + B_n(pr) n(n+1) Y_n], \\ \bar{\sigma}_\theta &= \mu \{E_n(pr) S_n + F_n(pr) n(n+1) Y_n + [G_n(pr) S_n + H_n(pr) Y_n]_{,\theta\theta} + \\ &\quad + [r F_n(pr) W_n \operatorname{cosec} \theta]_{,\theta\phi}\}, \\ \bar{\sigma}_\phi &= \mu \{E_n(pr) S_n + F_n(pr) n(n+1) Y_n + (\operatorname{cosec}^2 \theta \partial^2 / \partial \phi^2 + \\ &\quad + \cot \theta \partial / \partial \theta) [G_n(pr) S_n + H_n(pr) Y_n] - [r F_n(pr) W_n \operatorname{cosec} \theta]_{,\theta\phi}\}, \\ \bar{\tau}_{r\theta} &= \mu \{[C_n(pr) S_n + D_n(pr) Y_n]_{,\theta} + [\frac{1}{2} r B_n(pr) W_n]_{,\phi} \operatorname{cosec} \theta\}, \\ \bar{\tau}_{r\phi} &= \mu \{[C_n(pr) S_n + D_n(pr) Y_n]_{,\phi} \operatorname{cosec} \theta - [\frac{1}{2} r B_n(pr) W_n]_{,\theta}\}, \\ \bar{\tau}_{\theta\phi} &= \mu \{[G_n(pr) S_n + H_n(pr) Y_n] \operatorname{cosec} \theta\}_{,\theta\phi} - \\ &\quad - \mu [\partial^2 / \partial \theta^2 + n(n+1)/2] [r F_n(pr) W_n],\end{aligned}\quad (25)$$

where

$$\begin{aligned}
 A_n(pr) &= r^{-\frac{1}{2}}(\alpha pr)^{-2}[(\beta pr)^2 - 2(n-1)(n+2)]H_{n+\frac{1}{2}}^{(1)}(\alpha pr) + \\
 &\quad + 4\alpha pr H_{n+\frac{1}{2}}^{(1)}(\alpha pr) - 6H_{n+\frac{1}{2}}^{(1)}(\alpha pr), \\
 B_n(pr) &= r^{-\frac{1}{2}}[2\beta pr H_{n+\frac{1}{2}}^{(1)}(\beta pr) - 3H_{n+\frac{1}{2}}^{(1)}(\beta pr)], \\
 C_n(pr) &= -r^{-\frac{1}{2}}(\alpha pr)^{-2}[2\alpha pr H_{n+\frac{1}{2}}^{(1)}(\alpha pr) - 3H_{n+\frac{1}{2}}^{(1)}(\alpha pr)], \\
 D_n(pr) &= -r^{-\frac{1}{2}}[(\beta pr)^2 - 2(n-1)(n+2)]H_{n+\frac{1}{2}}^{(1)}(\beta pr) + \\
 &\quad + 2\beta pr H_{n+\frac{1}{2}}^{(1)}(\beta pr) - 3H_{n+\frac{1}{2}}^{(1)}(\beta pr), \\
 E_n(pr) &= r^{-\frac{1}{2}}(\alpha pr)^{-2}[(\beta pr)^2 - 2(\alpha pr)^2 - 2]H_{n+\frac{1}{2}}^{(1)}(\alpha pr) - \\
 &\quad - [2(\alpha pr)H_{n+\frac{1}{2}}^{(1)}(\alpha pr) - 3H_{n+\frac{1}{2}}^{(1)}(\alpha pr)], \\
 F_n(pr) &= 2r^{-\frac{1}{2}}H_{n+\frac{1}{2}}^{(1)}(\beta pr), \\
 G_n(pr) &= -2r^{-\frac{1}{2}}(\alpha pr)^{-2}H_{n+\frac{1}{2}}^{(1)}(\alpha pr), \\
 H_n(pr) &= r^{-\frac{1}{2}}[H_{n+\frac{1}{2}}^{(1)}(\beta pr) + 2\beta pr H_{n+\frac{1}{2}}^{(1)}(\beta pr)].
 \end{aligned}
 \tag{26}$$

Here primes denote differentiation with respect to the argument of the functions.

The Fourier transforms of the boundary conditions (4) give:

$$\begin{aligned}
 \bar{R} &= \mu \sum_n [A_n(pa)S_n + B_n(pa)n(n+1)Y_n], \\
 \bar{\Theta} &= \mu \sum_n \{[C_n(pa)S_n + D_n(pa)Y_n]_{,\theta} + [\frac{1}{2}aB_n(pa)W_n]_{,\phi} \operatorname{cosec} \theta\}, \\
 \bar{\Phi} &= \mu \sum_n \{[C_n(pa)S_n + D_n(pa)Y_n]_{,\phi} \operatorname{cosec} \theta - [\frac{1}{2}aB_n(pa)W_n]_{,\theta}\}.
 \end{aligned}
 \tag{27}$$

From these series we can determine the surface harmonics S_n , Y_n , and W_n . From (27) it is clear that a series representation of the following form is a convenient one:

$$\begin{aligned}
 \bar{R} &= \sum_n \bar{R}_n, \quad \bar{\Theta} = \sum_n (\bar{T}_{n,\theta} + \bar{\tau}_{n,\phi} \operatorname{cosec} \theta), \\
 \bar{\Phi} &= \sum_n (\bar{T}_{n,\phi} \operatorname{cosec} \theta - \bar{\tau}_{n,\theta}).
 \end{aligned}
 \tag{28}$$

This particular form is also seen to have the physical significance that the tangential tractions consist of components $\sum \bar{T}_{n,\theta}$ in θ -direction and $\sum \bar{T}_{n,\phi} \operatorname{cosec} \theta$ in ϕ -direction, which are derivable from a potential, and of components $\sum \bar{\tau}_{n,\phi} \operatorname{cosec} \theta$ in θ -direction and $\sum \bar{\tau}_{n,\theta}$ in ϕ -direction, which are derivable from a stream function. Here \bar{R}_n , \bar{T}_n , and $\bar{\tau}_n$ are surface spherical harmonics of integral degree n , of which \bar{R}_n is given by:

$$\bar{R}_n = \frac{1}{2}a_n^0 P_n(\cos \theta) + \sum_{m=1}^n (a_n^m \cos m\phi + b_n^m \sin m\phi) P_n^m(\cos \theta),
 \tag{29}$$

where a_n^m and b_n^m are functions of p . Surface harmonics \bar{T}_n and $\bar{\tau}_n$ have similar forms, with \bar{R}_n , a_n^m , and b_n^m replaced by \bar{T}_n , c_n^m , and d_n^m in the expression for \bar{T}_n , and they are replaced by $n(n+1)\bar{\tau}$, e_n^m , and f_n^m in the

expression for $\bar{\tau}_n$. Given \bar{R} , we can determine a_n^m, b_n^m from the Fourier-Legendre expansion theorem:

$$a_n^m = \frac{2n+1}{2\pi} \frac{(n-m)!}{(n+m)!} \int_{-\pi}^{\pi} \int_0^{\pi} \bar{R} P_n^m(\cos \theta) \cos m\phi \sin \theta \, d\theta d\phi, \quad (30)$$

$$b_n^m = \frac{2n+1}{2\pi} \frac{(n-m)!}{(n+m)!} \int_{-\pi}^{\pi} \int_0^{\pi} \bar{R} P_n^m(\cos \theta) \sin m\phi \sin \theta \, d\theta d\phi.$$

(26) In order to obtain \bar{T}_n and $\bar{\tau}_n$ we use the following differential equation satisfied by the spherical surface harmonics \bar{T}_n ,

$$(\sin \theta)^{-1} (\sin \theta \bar{T}_{n,\theta})_{,\theta} + (\sin \theta)^{-2} \bar{T}_{n,\phi\phi} = -n(n+1) \bar{T}_n. \quad (31)$$

This of course is also valid for $\bar{\tau}_n$. If we now form the left side of (31) from (28) we obtain

$$\begin{aligned} \sum n(n+1) \bar{T}_n &= -\operatorname{cosec} \theta [(\Theta \sin \theta)_{,\theta} + \Phi_{,\phi}] = \bar{T}, \\ \sum n(n+1) \bar{\tau}_n &= -\operatorname{cosec} \theta [\Theta_{,\phi} - (\Phi \sin \theta)_{,\theta}] = \bar{\tau}. \end{aligned} \quad (32)$$

These equations are of the same form as the first equation of (28). Hence using $\bar{T}/n(n+1)$ and $\bar{\tau}/n(n+1)$ with \bar{T} and $\bar{\tau}$ defined in terms of Θ and Φ by (32) in place of R in (30), we obtain the coefficients c_n^m, d_n^m , and e_n^m, f_n^m respectively.

(27) We now substitute (27) into the first equation of (28) and into (32), the latter being equivalent to the remaining two equations of (28). Using (31) for all surface harmonics we obtain

$$\begin{aligned} A_n(pa) S_n + B_n(pa) n(n+1) Y_n &= \mu^{-1} \bar{R}_n, \\ C_n(pa) S_n + D_n(pa) Y_n &= \mu^{-1} \bar{T}_n, \\ a B_n(pa) W_n &= 2\mu^{-1} \bar{\tau}_n. \end{aligned} \quad (33)$$

(28) The solution of (33) is

$$\begin{aligned} S_n &= (\mu \Delta_n)^{-1} [D_n(pa) \bar{R}_n - B_n(pa) n(n+1) \bar{T}_n], \\ Y_n &= (\mu \Delta_n)^{-1} [A_n(pa) \bar{T}_n - C_n(pa) \bar{R}_n], \\ W_n &= 2[a\mu B_n(pa)]^{-1} \bar{\tau}_n, \end{aligned} \quad (34)$$

where

$$\Delta_n \equiv \Delta_n(pa) = A_n(pa) D_n(pa) - n(n+1) B_n(pa) C_n(pa). \quad (35)$$

(29) This of course completes the problem, for (34) used in (24) and (25) gives the Fourier transforms of displacement and stress components; on multiplication by $(2\pi)^{-1} e^{-ipx}$ and integrating with respect to p from $-\infty$ to ∞ we obtain the actual displacement and stress components.

We now give some special cases which are of practical interest.

4. Special cases

- (i) *Zero surface shear.* $\Theta = \Phi = 0$. Hence $\bar{T}_n = \bar{\tau}_n = 0$.
 (ii) *Normal traction with central symmetry and (i).* In this case $a_n^m = b_n^m = 0$ when m is odd.
 (iii) *Normal traction with constant amplitude over a cap* $0 \leq \theta \leq \theta_1$, $-\pi \leq \phi \leq \pi$ and (i), i.e.

$$\bar{R}(\theta, \phi, p) = \begin{cases} \bar{R}(p) & \text{when } 0 \leq \theta \leq \theta_1, -\pi \leq \phi \leq \pi \\ 0 & \text{otherwise.} \end{cases} \quad (36)$$

Substituting (36) into (30) gives:

$$b_n^m = 0, \quad a_n^0 = (2n+1)\bar{R}(p)P_n^{-1}(\cos \theta_1)\sin \theta_1, \quad a_n^m = 0 \quad (m \neq 0). \quad (37)$$

- (iv) *Normal traction with constant amplitude over a zone* $\theta_0 \leq \theta \leq \theta_1$, $-\pi \leq \phi \leq \pi$, and (i). This follows from (37) by substituting θ_0 for θ_1 in (37) and subtracting the result from (37). Hence

$$b_n^m = 0, \quad a_n^0 = (2n+1)\bar{R}(p)[P_n^{-1}(\cos \theta_1)\sin \theta_1 - P_n^{-1}(\cos \theta_0)\sin \theta_0], \\ a_n^m = 0 \quad (m \neq 0). \quad (38)$$

- (v) *Concentrated load at one pole and (i).* In this case we have (37) with

$$\lim_{\substack{\theta_1 \rightarrow 0 \\ R \rightarrow \infty}} \bar{R}(p)2\pi a^2(1 - \cos \theta_1) = \bar{R}_0(p), \quad (39)$$

where $\bar{R}_0(p)$ is the Fourier transform of the concentrated time dependent load $R_0(t)$ acting at the pole $r = a$, $\theta = 0$. Using (37) in (39) and carrying out the indicated limit we obtain:

$$b_n^m = 0, \quad a_n^0 = \frac{2n+1}{4\pi a^2} \bar{R}_0(p), \quad a_n^m = 0 \quad (m \neq 0). \quad (40)$$

- (vi) *Impulsive concentrated load and (i).* In this case we write $R_0(t) = R_0 \delta(t)$ where R_0 is constant and $\delta(t)$ is the Dirac delta function defined by

$$\delta(t) = \begin{cases} \infty & \text{for } t = 0 \\ 0 & \text{for } t \neq 0; \end{cases} \quad \int_{-\infty}^{\infty} \delta(t) dt = 1.$$

Hence
$$b_n^m = 0, \quad a_n^0 = \frac{2n+1}{4\pi a^2} R_0, \quad a_n^m = 0 \quad (m \neq 0). \quad (41)$$

- (vii) *Concentrated ring load over a circle* $\theta = \theta_1$, $-\pi \leq \phi \leq \pi$ and (i). This is a limiting case obtained from (38) by letting

$$\lim_{\substack{\theta_1 \rightarrow \theta_0 \\ R \rightarrow \infty}} \bar{R}(p)2\pi a^2(\cos \theta_0 - \cos \theta_1) = \bar{R}_0(p). \quad (42)$$

This gives (40) with total load $R_0(t)$. The load per unit length of the circle is $R_0(t)/2\pi a \sin \theta_1$.

The case of *impulsive concentrated ring load* follows from this, and

results in (42) with load per unit length $R_0/2\pi a \sin \theta_1$, where R_0 has the same meaning as in (vi).

(viii) *Uniform tractions and* (i). If we take $\theta_1 = \pi$ in (iii) we obtain the case of uniform time dependent tractions applied to the cavity. Taking the limit as $\theta_1 \rightarrow \pi$ in (37) or carrying out the integration in (30) we find

$$b_n^m = 0, \quad a_0^0 = 2\bar{R}(p), \quad a_n^m = 0 \quad (m, n \neq 0). \quad (43)$$

Components of displacements and stress follow from (24) and (25):

$$\begin{aligned} -4\alpha\mu u &= \frac{1}{2\pi x^2} \int_{-\infty}^{\infty} \bar{R}(\xi/a\alpha) \frac{\xi x + i}{-i(\gamma\xi)^2 + \xi + i} e^{i\xi(x-1-y)} d\xi, \\ \alpha\alpha\sigma_r &= \frac{1}{2\pi x^3} \int_{-\infty}^{\infty} \bar{R}(\xi/a\alpha) \frac{-i(\gamma\xi x)^2 + \xi x + i}{-i(\gamma\xi)^2 + \xi + i} e^{i\xi(x-1-y)} d\xi, \\ \alpha\alpha\sigma_\theta &= \frac{1}{4\pi x^3} \int_{-\infty}^{\infty} \bar{R}(\xi/a\alpha) \frac{i(1-2\gamma^2)(\xi x)^2 - \xi x - i}{-i(\gamma\xi)^2 + \xi + i} e^{i\xi(x-1-y)} d\xi. \end{aligned} \quad (44)$$

$$v = w = \tau_{r\theta} = \tau_{r\phi} = \tau_{\theta\phi} = 0, \quad \sigma_\phi = \sigma_\theta,$$

$$x = r/a, \quad y = t/a\alpha, \quad \gamma = \beta/2\alpha.$$

Using the convolution theorem and noticing that the inverse Fourier transform of $\bar{R}(\xi/a\alpha)$ is $a\alpha R(a\alpha y)$, the first integral in (44) may be evaluated. To this end we note that the coefficient of $\bar{R}(\xi/a\alpha)$ has two simple poles, both of which are located in the lower half of the ξ -plane. Hence a contour consisting of the real axis and a half circle in the lower half plane containing both poles with centre at the origin may be chosen to determine the inverse Fourier transforms of the fraction in the integrand of u . The convolution theorem then gives u . Using this result in (2) and (3) we obtain σ_r and σ_θ .

$$u = \frac{1}{2}a\mu^{-1}(4\gamma^2-1)^{-\frac{1}{2}} \operatorname{re} x^{-2}(-sx+i)R^*(y+1-x),$$

$$\begin{aligned} \sigma_r &= x^{-1}R[a\alpha(y+1-x)] + \\ &\quad + 2(4\gamma^2-1)^{-\frac{1}{2}}x^{-3} \operatorname{re} [i(\gamma sx)^2 + sx - i]R^*(y+1-x), \end{aligned} \quad (45)$$

$$\begin{aligned} \sigma_\theta &= (1-\frac{1}{2}\gamma^2)x^{-1}R[a\alpha(y+1-x)] + \\ &\quad + (4\gamma^2-1)^{-\frac{1}{2}}x^{-3} \operatorname{re} [i(2\gamma^2-1)(sx)^2 - sx + i]R^*(y+1-x), \end{aligned}$$

where

$$R^*(y+1-x) = \int_{-\infty}^{y+1-x} R(a\alpha\eta) e^{is(y+1-x-\eta)} d\eta$$

valid for $y > x-1$. For $y < x-1$ all quantities are zero. If the function $R(t)$ is discontinuous at $t = 0$ we define $R(t)$ at $t = 0$ by

$$R(t) = \begin{cases} 0 & \text{for } t = 0- \\ \frac{1}{2}R & \text{for } t = 0 \\ R & \text{for } t = 0+. \end{cases}$$

In return, when any of u , σ_r , σ_θ is discontinuous at $y+1-x = 0$ Fourier's theorem gives one-half of the values obtainable from (45) at $y+1-x = 0$. Physically $y = x-1$ represents the wave front. Naturally at those points ahead of the wave front the stress and the displacement components must vanish, as the analysis predicts, while at the points behind the wave front they are given by (45). On substituting the values of x and y from (44) we express the wave front by $t/\alpha = r-a$. Here $1/\alpha$ is the velocity of the dilatational waves. In this particular case, we only have the dilatational waves. This of course is also indicated by the type of boundary loading on the surface $r = a$.

(ix) *Uniform pulsating load and* (i). This type of normal tractions may be expressed by

$$R(\theta, \phi, t) = R_0 e^{-i\omega t}. \quad (46)$$

Hence

$$\bar{R}(\theta, \phi, p) = 2\pi R_0 \delta(\omega - p),$$

$$b_n^m = 0, \quad \alpha_0^0 = 4\pi R_0 \delta(\omega - p), \quad \alpha_n^m = 0 \quad (m, n \neq 0). \quad (47)$$

In obtaining (47) we used the formal relation

$$2\pi\delta(p) = \int_{-\infty}^{\infty} e^{-ipt} dt.$$

Substituting the value of \bar{R} given by (47) into (44) we obtain

$$\begin{aligned} 4\mu u/aR_0 &= x^{-2}(\Omega x + i)[i(\gamma\Omega)^2 - \Omega - i]^{-1} \exp[i\Omega(x-1-y)], \\ \sigma_r/R_0 &= x^{-3}[i(\gamma\Omega x)^2 - \Omega x - i][i(\gamma\Omega)^2 - \Omega - i]^{-1} \exp[i\Omega(x-1-y)], \\ 2\sigma_\theta/R_0 &= x^{-3}[i(2\gamma^2 - 1)(\Omega x)^2 + \Omega x + i][i(\gamma\Omega)^2 - \Omega - i]^{-1} \exp[i\Omega(x-1-y)], \\ v = w = \tau_{r\theta} = \tau_{r\phi} = \tau_{\theta\phi} &= 0, \quad \Omega = a\alpha\omega, \quad y > x-1. \end{aligned} \quad (48)$$

(x) *Blast loading and* (i). In this case we have $R = -R_0 \delta(t)$. This is a special case of (viii) in which $\bar{R}(p) = -R_0$. The substitution of

$$R(a\alpha\eta) = -R_0 \delta(a\alpha\eta) = -R_0 \delta(\eta)/a\alpha$$

in (45) gives the displacement and stress components,

$$\begin{aligned}
 2\mu x \left(\frac{2}{\Gamma} - 1 \right)^{\frac{1}{2}} u / R_0 &= x^{-2} \{ (1 - \Gamma x) \sin[(2\Gamma - \Gamma^2)^{\frac{1}{2}}(y + 1 - x)] + \\
 &\quad + x(2\Gamma - \Gamma^2)^{\frac{1}{2}} \cos[(2\Gamma - \Gamma^2)^{\frac{1}{2}}(y + 1 - x)] \} e^{-\Gamma(y+1-x)}, \\
 \left(\frac{2}{\Gamma} - 1 \right)^{\frac{1}{2}} \alpha a \sigma_r / 2R_0 &= -\frac{1}{2} \left(\frac{2}{\Gamma} - 1 \right)^{\frac{1}{2}} x^{-1} \delta(y + 1 - x) + \\
 &\quad + x^{-3} \{ [(1 - \Gamma)x^2 + \Gamma x - 1] \sin[(2\Gamma - \Gamma^2)^{\frac{1}{2}}(y + 1 - x)] + \\
 &\quad + (2\Gamma - \Gamma^2)^{\frac{1}{2}}(x^2 - x) \cos[(2\Gamma - \Gamma^2)^{\frac{1}{2}}(y + 1 - x)] \} e^{-\Gamma(y+1-x)}, \\
 \left(\frac{2}{\Gamma} - 1 \right)^{\frac{1}{2}} \alpha a \sigma_\theta / R_0 &= -\left(\frac{2}{\Gamma} - 1 \right)^{\frac{1}{2}} (1 - \Gamma)x^{-1} \delta(y + 1 - x) + \\
 &\quad + x^{-3} \{ [2(1 - \Gamma)^2 x^2 - \Gamma x + 1] \sin[(2\Gamma - \Gamma^2)^{\frac{1}{2}}(y + 1 - x)] + \\
 &\quad + (2\Gamma - \Gamma^2)^{\frac{1}{2}} [2(1 - \Gamma)x^2 + x] \cos[(2\Gamma - \Gamma^2)^{\frac{1}{2}}(y + 1 - x)] \} e^{-\Gamma(y+1-x)}, \\
 \sigma_\phi &= \sigma_\theta, \quad v = w = \tau_{r\theta} = \tau_{r\phi} = \tau_{\theta\phi} = 0, \quad \Gamma = \frac{1}{2}\gamma^{-2},
 \end{aligned} \tag{49}$$

valid for $y > x - 1$. For the case of $y < x - 1$ all quantities are zero. At the wave front $y + 1 - x = 0$ one-half of the values obtained from (49) by substitution of $y + 1 - x = 0$ should be understood. For example at $t = 0$ we have:

$$u(1, 0) = \begin{cases} 0 & \text{for } t = 0- \\ \frac{1}{2} [\rho(\lambda + 2\mu)]^{-\frac{1}{2}} R_0 & \text{for } t = 0 \\ [\rho(\lambda + 2\mu)]^{-\frac{1}{2}} R_0 & \text{for } t = 0+. \end{cases} \tag{50}$$

As indicated by Fig. 2 the value of u at $x = 1, t = 0+$ is the one to which $u(1, y)$ converges as $y \rightarrow 0$.

Finally, it may be of interest to note that the problems dealing with spherically symmetric type of loading such as (i), (vi), (viii), (ix), and (x) can be treated more simply by solving the one-dimensional problem involving a single partial differential equation for u alone. This was done for the cases (viii) and (x) and the results were the same as in this paper.

5. Computation and discussion

The resulting formulae (49) in the case of blast type of loading were used to compute the radial displacement u and the stress components σ_r and σ_θ . The computations were carried out on an electronic digital computer. The results are plotted on Figs. 2-4 versus the non-dimensional radial distance x for values of non-dimensional time $y = \frac{1}{4}, \frac{1}{2}, 1, 2, 3, \dots, 11$ as a parameter and for two distinct Poisson's ratios $\nu = \frac{1}{3}$ and $\frac{1}{4}$. Here $y = t/\alpha x$ is the ratio of the distance that the dilatational waves travel in time t , to the radius a of the sphere. Some interesting results which follow from these figures may be noted. At $t = 0$ a blast of impulse R_0 is applied to $x = 1$, the surface of

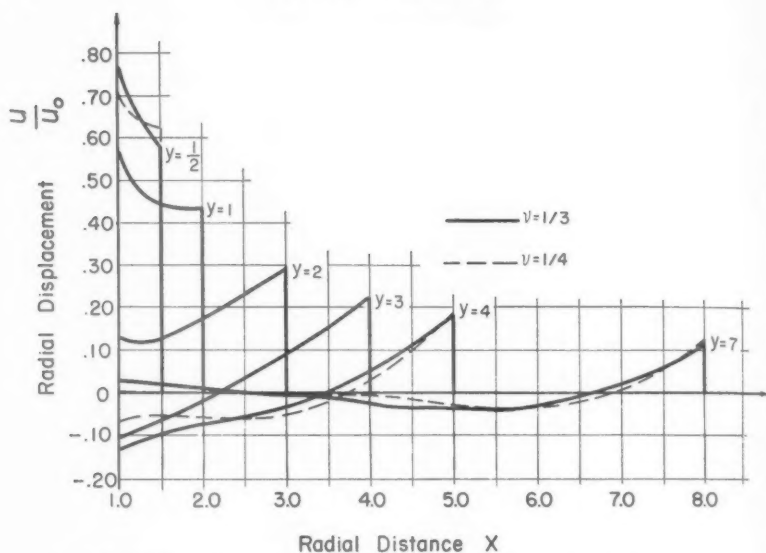


FIG. 2. Radial displacement produced by a blast. $u_0 = R_0/2\mu\alpha(2/\Gamma-1)^{1/2}$

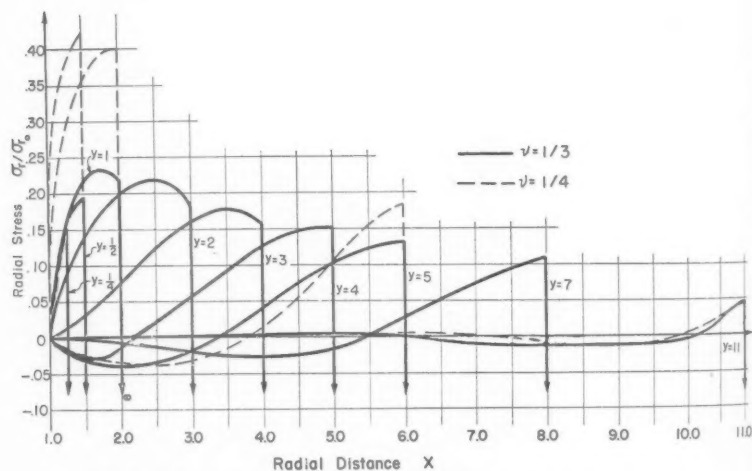


FIG. 3. Radial stress produced by a blast. $\sigma_{r_0} = 2R_0/\alpha a(2/\Gamma-1)^{1/2}$

the cavity. Immediately thereafter we find that a positive displacement $u = u_0$ takes place. As time passes the displacement waves move into the media with sharp fronts ahead of which everything is zero. The variation of displacements in space and in time follows the interesting pattern shown

in Fig. 2. We notice, however, that after a while the tail of the wave front becomes negative, thus indicating negative displacements at some points behind but becoming positive again later at the same points. The pattern of the hoop stress σ_θ (Fig. 4) is similar to this except that after a sufficient lapse of time we find the maximum σ_θ not at the wave front but slightly behind it. At the beginning σ_θ is tensile, but later it becomes compressive and then tensile.

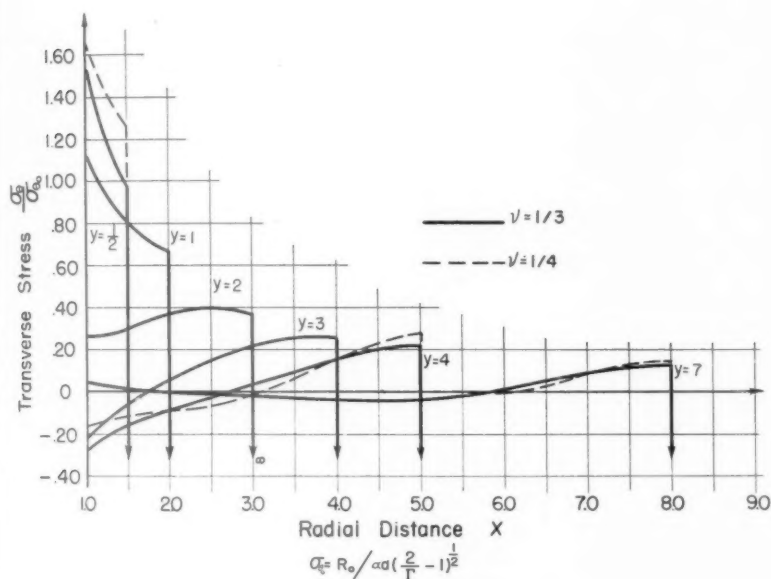


FIG. 4. Hoop stress produced by a blast

Perhaps the most unexpected result is that the radial stress in the neighbourhood of the spherical cavity immediately after the application of blast is tensile. It starts with the zero value at $x = 1$ and grows into a maximum which is slightly behind the wave front. The mathematical formulae at the wave front $y + 1 - x = 0$ should be interpreted as explained above. At the wave front we find infinite compressive stress. The study of the case of rectangular pulse in the limiting case also indicates this fact which is no more than the propagating blast pulse of finite impulse applied to the surface of the cavity.

The author is indebted to Mr. J. C. Samuels for checking the analysis and for carrying out the computations.

REFERENCES

1. K. SEZAWA, *Bull. Earthq. Res. Inst. Tokyo Imperial U.* **13** (1935) 740; **14** (1936) 506.
2. ——— and K. KANAI, *ibid.* **14** (1936) 10; **15** (1937) 13.
3. W. INOUE, *ibid.* **14** (1936) 582; **15** (1937) 90; **15** (1937) 956.
4. H. KAWASUMI and R. YOSIYAMA, *Disin.* **7** (1935) 359.
5. G. NISHIMURA, *Bull. Earthq. Res. Inst. Tokyo Imperial U.* **15** (1937) 614.
6. J. VANĚK, *Czech. J. Phys.* **3** (1953) 97.
7. W. A. ALLEN and W. GOLDSMITH, *J. App. Phys.* **26** (1955) 69.
8. W. GOLDSMITH and W. A. ALLEN, *J. Acoust. Soc. Amer.* **27** (1955) 47.
9. J. A. SHARPE, *Geophysics*, **7** (1942) 144, 311.
10. F. G. BLAKE, JR., *J. Acoust. Soc. Amer.* **24** (1952) 211.

THE INDENTATION OF A THICK SHEET OF ELASTIC MATERIAL BY A RIGID CYLINDER

By DAPHNE G. PADFIELD and JEAN SIDA

(Wool Industries Research Association)

[Received 15 November 1956]

SUMMARY

The indentation of a uniform elastic sheet by a flat-ended cylinder has been computed for different ratios of cylinder radius to depth of sheet. The method used is a direction application of the Hankel transform method described by Sneddon (1).

1. Description of problem and method

THE indentation of the plane surface of a semi-infinite elastic medium by a rigid punch in the shape of a flat-ended cylinder has been fully discussed by Sneddon (1, p. 458). In this note we give an approximate solution of the corresponding problem for a thick elastic sheet. We envisage a uniform elastic sheet resting on a smooth rigid horizontal table, and wish to find the depth and the shape of the indentation caused when a heavy flat-ended circular cylinder is placed on top of the sheet (Fig. 1 (i)). The same analysis will also apply to the situation shown in Fig. 1 (ii) where the free sheet of material is acted upon by opposed loads on symmetrically placed rigid circular faces. The appropriate boundary conditions on the indented surface are: the depression constant within the circle of contact with the cylinder face, and the surface pressure zero outside this circle.

From the theory given in (1), pp. 468 ff. it is easy to calculate the surface depression for any known pressure distribution, the numerical work being light when the function giving the pressure distribution has a known analytical Hankel transform.† Three distributions of pressure $p(r)$ over a circular area (i.e. zero outside the circle) are particularly convenient; these are distributions proportional to

$$\left. \begin{array}{ll} \text{(i)} & 1 \\ \text{(ii)} & a^2 - r^2 \\ \text{(iii)} & (a^2 - r^2)^{-\frac{1}{2}} \end{array} \right\} (0 < r < a), \text{ cf. (1), p. 528,}$$

where a is the radius of the circle. For each of these pressure distributions

† The Hankel transform $p(r)$ is defined by $\bar{p}(\xi) = \int_0^\infty r p(r) J_n(\xi r) dr$. The inverse transform relation (assuming convergence, etc.) is symmetrical, viz. $p(r) = \int_0^\infty \xi \bar{p}(\xi) J_n(\xi r) d\xi$. In the present work all transforms are for $n = 0$.

we may readily work out the corresponding depression. Thus the depression is

$$u(r) = \frac{2(1-\nu^2)}{E} \int_0^\infty \frac{\sinh^2 \xi d J_0(\xi r)}{\sinh \xi d \cosh \xi d + \xi d} \bar{p}(\xi) d\xi, \quad (1)$$

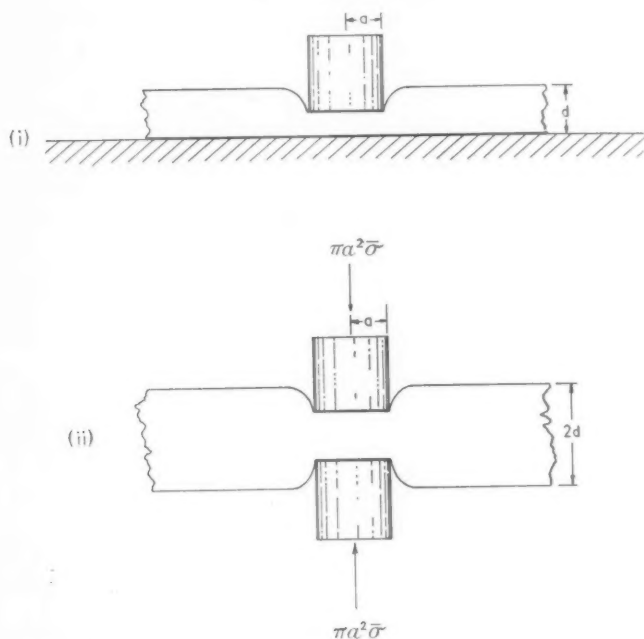


FIG. 1. (i) Indentation of thick sheet on table by heavy cylinder. (ii) Indentation of free sheet by opposed forces on rigid circular faces.

where d is the depth of the sheet, and $\bar{p}(\xi)$ takes the forms

- (i) $a J_1(a\xi)/\xi$,
- (ii) $4a J_1(a\xi)/\xi^3 - 2a^2 J_0(a\xi)/\xi^2$,
- (iii) $\sin(a\xi)/\xi$,

respectively. The symbols ν , E in equation (1) denote Poisson's ratio and Young's modulus.

By taking a linear combination of the three cases (i), (ii), and (iii), $u(r) = Au_1(r) + Bu_2(r) + Cu_3(r)$, say, we can approximate to the required boundary conditions, viz.

$$u(r) = \text{constant} = u, \text{ say } (0 < r < a),$$

$$\text{surface pressure} = 0 \quad (a < r).$$

In the actual computations three values of a/d were considered, viz.

$a/d = 1, 2, 4$. These results were supplemented by special solutions for the extreme cases $a/d = 0, \infty$ (see section 2). For each value of a/d the depression $u(r)$ was evaluated numerically from equation (1) for $r/a = 0.0, 0.2, 0.4, 0.6, 0.8, 1.0$ as well as for a number of values greater than unity. Constants A, B, C were then chosen to minimize

$$\sum w(r)\{Au_1(r) + Bu_2(r) + Cu_3(r) - 1\}^2,$$

the summation being over the values $r = 0, 0.2a, 0.4a, \dots, a$, and the weighting factors $w(r)$ being proportional to the areas of the corresponding r -zones. Knowing A, B , and C the ratio of the depression to the total load was calculated, and the profile of the sheet for $r > a$ was found.

2. The extreme cases $a/d = 0, \infty$

The necessary information about the case $a/d = 0$ is contained in Sneddon's solution of the problem of indentation of a semi-infinite medium by a rigid cylinder. His result is that a total load $2Ea/(1-\nu^2)$ is required to produce a depression $u = 1$.

The case $a/d = \infty$ can be examined by letting $d \rightarrow 0$ in the expression for $u(r)$ in equation (1). Thus,

$$\frac{u(r)}{d} \rightarrow \frac{(1-\nu^2)}{E} \int_0^\infty \xi \bar{p}(\xi) J_0(\xi r) d\xi = \frac{(1-\nu^2)}{E} p(r).$$

This shows that for $0 < r < a$, $p(r) = \text{constant} = (1-\nu^2)^{-1}Eu/d$.

3. Results

Let $\bar{\sigma}$ denote the mean pressure applied by the circular base of the cylinder = total weight $\div \pi a^2$. Then $(u/d)/\bar{\sigma} = (1-\nu^2)h/E$ where, from section 2, we find that $h = 0$ for $a/d = 0$, and $h = 1$ for $a/d = \infty$. The results of the calculations described in section 1 show that

$$h = 0.716 \quad \text{for } a/d = 1,$$

$$h = 0.844 \quad \text{for } a/d = 2,$$

$$h = 0.903 \quad \text{for } a/d = 4.$$

The form of Sneddon's result also indicates that $h \sim \frac{1}{2}\pi a/d$ at $a/d = 0$, i.e. it shows that the slope of the h v. (a/d) curve at the origin is $\frac{1}{2}\pi$. This curve is consequently fairly well defined for all a/d and is shown in Fig. 2. The profiles of the depression for $a/d = 1, 2, 4$ are shown in Fig. 3.

The residues in the least squares calculations give an underestimate of the magnitude of the departure of the calculated depression from $u = 1$

($r \leq a$). The residues were:

$\frac{a/d}{r/a} =$	1	2	4
0.0	-0.0012	+0.0038	+0.0280
0.2	+0.0003	+0.0104	+0.0174
0.4	-0.0001	-0.0026	-0.0010
0.6	+0.0000	-0.0114	-0.0185
0.8	-0.0001	+0.0112	+0.0119
1.0	+0.0000	-0.0032	-0.0024

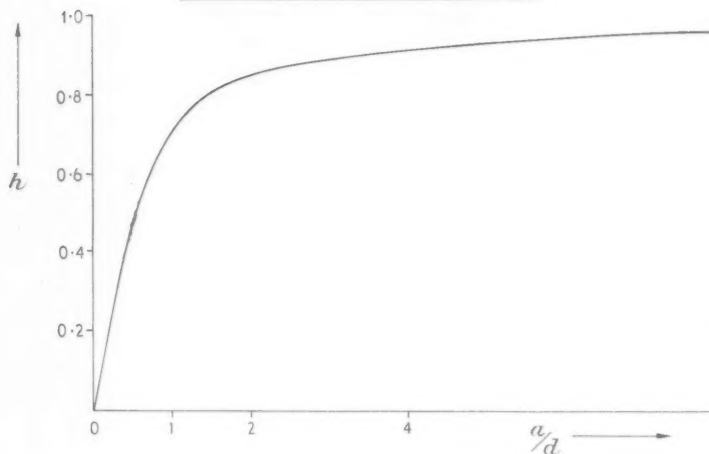


FIG. 2. Variation of h with a/d . $\left[\frac{u/d}{\partial} = \frac{h(1-\nu^*)}{E} \right]$

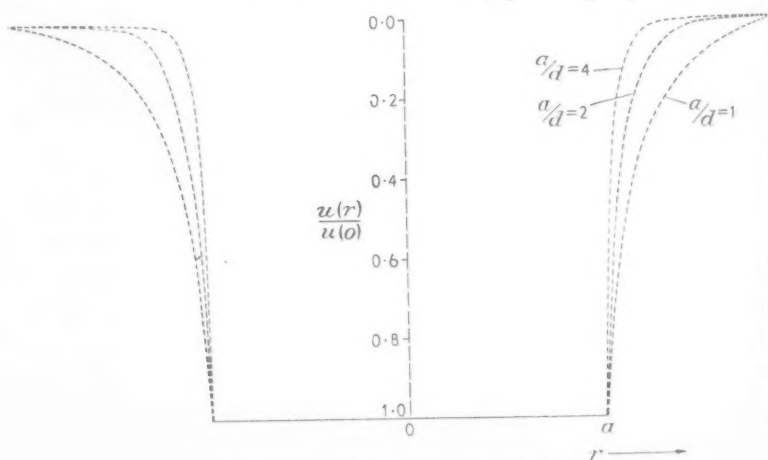


FIG. 3. Profile of depression for $a/d = 1, 2, 4$

The actual values of the coefficients A , B , C may also be of interest, since they determine the pressure distribution over the region of contact. The values were (in units of $\frac{1}{2}E/(1-\nu^2)$):

$a/d =$	1	2	4
$A =$	0.1676	0.7523	3.4921
$B =$	0.8854	0.8211	0.3446
$C =$	1.2057	1.5840	1.9879

Acknowledgements

We should like to thank a referee for a simplification of the argument in section 2 for the case $a/d \rightarrow \infty$, Mr. Anderson of the Testing Section of the Wool Industries Research Association whose inquiry prompted the piece of work, and the Director of Research and Council of the Association for permission to publish the results.

REFERENCE

1. I. N. SNEDDON, *Fourier Transforms*, chapter 10 (1951).

EXPERIMENTS ON THE SLOW SWIRLING FLOW OF A VISCOUS LIQUID THROUGH A TUBE

By A. M. BINNIE (*Engineering Laboratory, Cambridge*)

[Received 13 September 1956]

SUMMARY

To investigate the nature of the flow pattern when swirling water passes slowly through a long, straight, and transparent tube, a short length at the middle of the tube was caused to revolve, and through numerous holes in this portion the water was admitted under gravity. Colour injection was employed, and just clear of the rotating length three alternative régimes were seen: (I) downstream over the whole cross-section; (II) upstream near the axis and downstream near the wall; (III) downstream near the axis and the wall and upstream in the intermediate region. For four different arrangements of blockage in the two fixed lengths of tube, the results are plotted on diagrams having for their axes two dimensionless quantities that depend respectively on the tangential velocity at admission and on the mean axial velocity through the tube. When one of the fixed lengths was blocked close to the revolving part and colour was injected through the holes nearest to the blockage, the result under some conditions was three streams, symmetrically disposed 120° apart, that passed through the other fixed length.

1. Introduction

WHEN swirling water passes through a convergent nozzle the axial component of velocity in part of the nozzle is sometimes reversed, and the water in that region returns into the reservoir from which it came. This effect, which was described by Binnie and Teare (1), clearly requires further investigation. It was obtained under complicated conditions. An air core existed, round which a boundary layer formed, having solid body rotation and an axial velocity higher than the average over the whole cross-section. Moreover, a reservoir was essential in which the water from the external supply pipes could be steadied; consequently the velocity distribution at the nozzle entrance was unknown.

To simplify the problem as far as possible, the nozzle was replaced by a long transparent tube of uniform diameter. A short length of the tube was caused to revolve about its axis, and water was admitted through holes in this portion, thus the swirl (tangential velocity) at entrance was known. To conform with the theoretical case that Miss G. Vaisey is examining by means of relaxation methods, the apparatus was made symmetrical, the perforated drum being at the middle of the tube. The velocities were kept very low, hence turbulence was avoided and an air core never formed.

In an arrangement of this kind, at least three flow régimes can occur in the fixed tube near the drum:

(I) With the drum stationary, the water moves downstream at every point in the cross-section.

(II) At the other extreme, with rotation of the drum but no discharge, a circulation is set up dividing the cross-section into two parts. Over the central zone the axial velocity is upstream, and downstream motion is confined to the annular region extending to the wall. It should be noted that the terms upstream and downstream applied to neighbouring currents do not imply a large shearing force between them, for both possess swirling motion in the same angular direction.

(III) Under certain conditions a treble division occurs; the water near the walls and in the region of the axis moves downstream while upstream flow takes place in the intermediate zone. This effect, which was described qualitatively by Nuttall (2), appears to resemble reverse flow in a nozzle.

The ranges of régimes I and II are not as small as has been indicated above. To find the limits of their existence and the conditions under which régime III occurs, four sets of experiments were carried out in turn with different arrangements of blockage in the fixed tubing:

- (A) Both the fixed lengths of tube were left unobstructed.
- (B) The outlet of the left-hand tube was blocked with a rubber bung.
- (C) A sliding bung with a central hole was placed in both tubes at a distance of four diameters from the drum.
- (D) The left-hand tube was blocked by a sliding bung fixed as near the drum as possible, and in the right-hand tube a sliding bung with a central hole was placed at a distance of four diameters from the drum.

The numerical results are presented in the form of values of two non-dimensional parameters, which are based respectively on the mean axial velocity in the tube and on the swirl at admission. These parameters are defined by

$$R = \frac{aW}{\nu} \quad \text{and} \quad V = \frac{aV}{\nu},$$

where a is the pipe radius, ν is the kinematic viscosity, W is the mean axial velocity of the water in the tube, and V is the swirl velocity imposed on it at admission. The first parameter corresponds to the Reynolds number of flow without swirl.

2. Description of apparatus

Perspex tubing of internal diameter 2 in. was employed, its total length being 66 in., the central 4 in. of which formed the revolving drum. The

tubing was supported horizontally, as shown in Fig. 1, in an open tank 10 in. wide and 18 in. deep, which was divided by two diaphragms into a central compartment 50 in. long and two end compartments each 23 in. long. The drum was held in place by three wheels at each end, and water tightness between it and the adjoining fixed lengths was maintained by Gaco rings lubricated by means of screw-down grease cups (not shown). It was rotated by an electric motor and a variable-speed gear box placed outside one end of the tank, the drive being transmitted through a long horizontal shaft to a small gear-wheel which engaged with a toothed ring on the drum. Water was led under gravity from a constant-level tank to a submerged inlet in the central compartment. After passing into the drum and dividing into equal streams in the two fixed lengths of tube, it reached the end compartments, which were connected by external piping to a small tank and weir. The flow through the apparatus was controlled by altering the heights of the constant-level tank and of the weir, and it was easily measured with a graduated flask held under the weir for a known time. The tests were usually carried out with the tube in the central compartment submerged a couple of inches, and even at the highest discharges the outlets of the fixed tubes were well covered. The perforations in the drum consisted of 11 belts, each of 60 holes 0.080 in. in diameter, symmetrically disposed 0.25 in. apart. The total cross-section of the holes was about 5 per cent greater than the cross-section of the tube. The wall thickness of the drum was 0.25 in., so that it is reasonable to take the swirl velocity imposed on the water at admission as equal to the tangential velocity of the inner surface of the drum. The revolutions were timed with a stop-watch, and the maximum speed that could be obtained without undesirable vibration was about 40 rev/min.

In order that the flow pattern inside the tube might be examined, three tapplings designated R_1 , R_2 , R_3 , were drilled along the top of the right-hand fixed tube at distances 2.65, 5.65, 8.65 in. from the central plane of the drum, and likewise three tapplings L_1 , L_2 , L_3 , were added symmetrically to the left-hand tube. At the tapplings, coloured liquid was slowly injected through a probe of outside diameter 0.043 in. mounted on a vertical cathetometer. The tip of the probe could be traversed across the diameter of the tube but, to avoid unnecessary disturbance, only the upper radius was usually explored. The coloured liquid was a solution in water of the sodium salt of eosin, made up to density 1.003 and then diluted with methyl alcohol to density unity.

The apparatus thus differed from that used by Talbot (3), who employed a long unperforated tube in which swirl was induced by the rotation of part of the tube situated some distance from the entry.

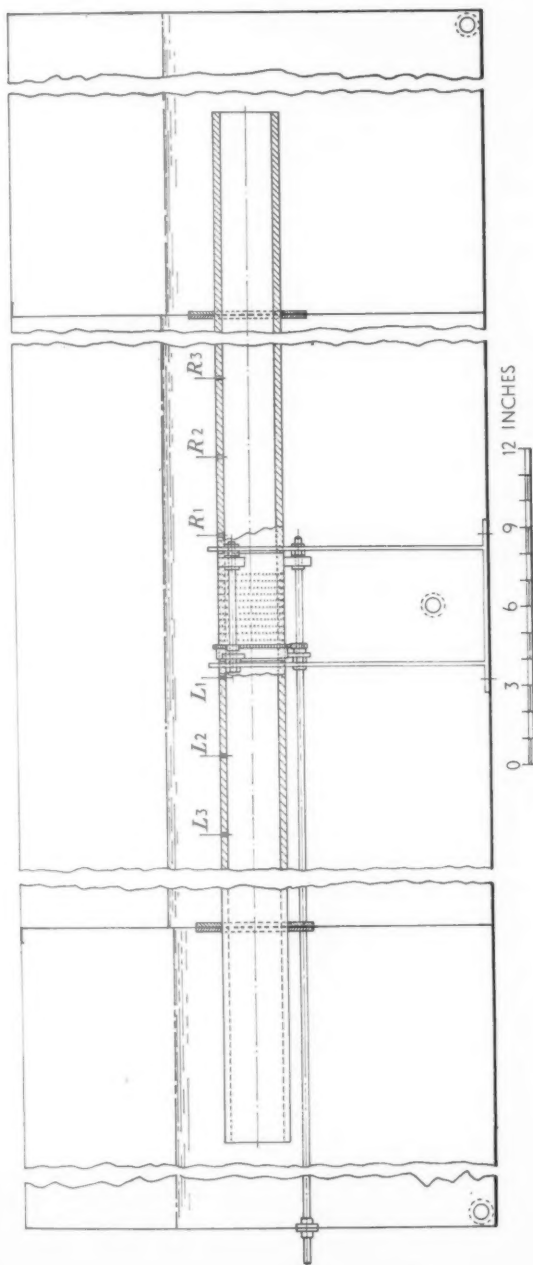


FIG. 1. Side elevation of apparatus

3. Experiments on symmetrical flow with no obstructions in the fixed tubes (arrangement A)

Preliminary tests first with no holes in the drum and then with a single row of 30 holes 0.020 in. in diameter demonstrated the existence of régime II and the necessity for a great number of holes in order to avoid turbulence at entry. Accordingly the 660 holes mentioned in section 2 were drilled; for reference the eleven rows of holes will be numbered from -5 through 0 to $+5$ reading from the left. To check the symmetry of the flow a bung was inserted at each outlet of the fixed tubes in turn, and with the outlet weir unaltered the discharge through the two halves of the apparatus was found to be the same. Colour was then inserted into the middle compartment outside but very close to the central row of holes, and after admission it was seen to divide symmetrically into two streams. The experiments of Binnie and Teare (1) with a double-ended swirl chamber operated at high swirls showed large axial velocities across the central plane, and a similar effect in the present apparatus was thought possible. But it was only on rare occasions that colour injected at R_1 found its way into the left-hand tube. In these few instances colour under the influence of régime II was carried up near the axis to the central plane where it remained visible for a matter of minutes, and occasionally a small proportion of the colour drifted from this stagnant zone across the central plane.

The usual method of test was to set the discharge and then to increase the speed of rotation in steps. At each step a measurement of the discharge was made which, owing to centrifugal action, decreased slightly with rising speed, and at R_1 colour was injected at $r/a = 0.9, 0.8, \dots, 0$, where r is the radius at which the tip of the probe was placed. The kind of régime was noted and the result plotted in Fig. 2. The two sets of observations at values of R about 140 and 260 were made when the outlet of each tube was obstructed by a bung with a $\frac{1}{4}$ -in. hole; the results are seen to be in agreement with the tests under normal conditions, and this was found to be true also of a set at about $R = 76$. At $V = 0$ and the largest values of R , upstream flow was found very close to the wall. This was part of a local eddy induced by the high velocity of the water at admission. It was thought reasonable to plot this observation as régime I, which was found at all lower values of R with the drum stationary.

To explain the nature of régime II an account will be given of what was seen at $V = 460$ (8 rev/min), $R = 104$ ($W = 11.3$ in./min). Under these conditions the coloured streamline from a fixed position of the probe did not constantly follow exactly the same path, but its wanderings did not obscure the general flow pattern. With injection at R_1 , the colour inserted

at $r/a = 0.9$ was carried downstream with vigorous swirl and axial velocities; both components greatly diminished beyond R_3 but the swirl was still perceptible as the colour entered the end compartment. It should be noted that inevitably the colour left the probe with a small radial velocity, therefore the effective radii of injection were somewhat less than the values given

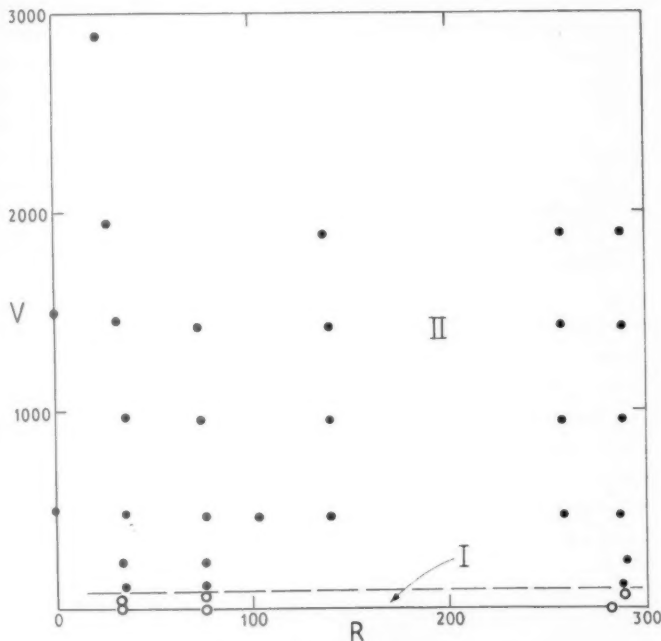


FIG. 2. Régime diagram for arrangement A

○ Régime I
● „ II

here. At $r/a = 0.8$ and 0.7 the initial pitch of the helix was smaller than before, but it rapidly increased downstream as the swirl died away. At $r/a = 0.6$ the initial axial velocity was zero; some of the colour was whirled away downstream and the rest moved upstream shedding filaments which, after moving to a larger radius, were caught in the fast downstream flow. Some succeeded in moving as far as row +5 before being turned back. At smaller values of r/a all the colour initially moved upstream, the shedding process in the fixed tube becoming less marked. At $r/a = 0.3$ all the colour entered the drum, in which its swirl and upstream velocities were very small until it drifted to a larger radius and was suddenly caught in the main stream. At values of $r/a < 0.3$ the colour reached the central plane but with extreme

slowness in the last part of its travel; from there a trace escaped into the left-hand tube. At $r/a = 0$ the colour on the central plane concentrated into a disk having a radius nearly half that of the tube. From the periphery traces were carried away downstream by the main flow, and a small proportion was removed into the left-hand tube in a similar manner.

These observations were confirmed by a traverse at R_3 . The position of zero axial velocity was at about $r/a = 0.45$, showing that the area of the core of upstream flow was diminished by a half in the length between R_1 and R_3 . External injections were also made in close proximity to rows 0, +2, and +5. From 0 the colour divided symmetrically and passed downstream at a little distance from the wall; but when the drum was clear and a minute had elapsed after the end of the injection, traces returned simultaneously into the drum from both fixed tubes at a radius of about $\frac{1}{2}$ in. From row +2 most went swirling away past R_3 but a little, after reaching a point between R_1 and R_2 , returned slowly at a smaller radius into the drum. From +5 the same reverse trace was seen, but the main flow was downstream so close to the wall that beyond R_3 most of its swirl was lost.

Almost the whole of Fig. 2 is occupied by régime II. Along the horizontal axis of the diagram régime I is bound to exist, but a very small amount of rotation sufficed to remove it, consequently it is confined to a narrow band at the bottom. Régime III was not found within the ranges of V and R that were explored; and as this type of flow was the principal object of the investigation, an attempt, described in the next section, was made to produce it by destroying the symmetry of the apparatus.

4. Experiments on asymmetrical flow produced by blocking the outlet of the left-hand tube (arrangement B)

In this series of tests the outlet of the left-hand tube was closed by a rubber bung. The alteration had the desired result, for over a certain range of R shown in Fig. 3, régime I changed to III when the speed of rotation was sufficiently increased. With the usual procedure of injection at R_1 , an almost constant discharge, and a start with zero rotation, the decay of régime I as the speed was raised was indicated by a fall in the axial velocity at about half radius, and the velocity at this position was eventually reversed. As R was increased, a higher rotational speed was required to effect the transformation. Fig. 3 differs greatly from Fig. 2. The observations along the axes were the same as before, but régime II appears nowhere else on the diagram, for it was eliminated by a slow rotation and replaced by I at low axial velocities and by III at high.

More complete observations were made of régime III at $V = 1,790$ (28 rev/min), $R = 68$ ($W = 6.6$ in./min), under which conditions the

streamlines were steadier than in the case detailed in section 3. With R_1 in use, the colour injected at $r/a = 0.9$ and 0.8 was carried downstream out of the central compartment, losing its swirl before it had travelled far. At $r/a = 0.7$ it circled upstream $\frac{1}{2}$ in. and was then carried away at larger radius by the main downstream flow. This effect was more pronounced at

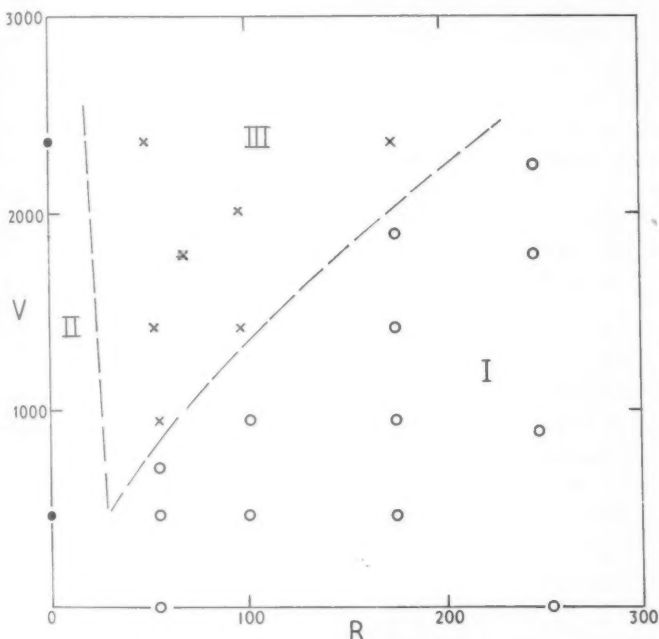


FIG. 3. Régime diagram for arrangement B

- Régime I
 ● " II
 × " III

radii down to $r/a = 0.4$, the colour getting just inside the drum before being swept downstream. For smaller radii the initial axial velocity was downstream, though at $r/a = 0.3$ the travel in that direction was slight, being succeeded by a very slow movement just into the drum at a larger radius. At $r/a = 0.2$ much the same was observed, but the colour drifted into a stagnant zone in the form of a conical frustum with its smaller end about $\frac{3}{4}$ in. downstream from R_1 and its larger end in the drum between row +5 and the nearby joint. At $r/a = 0$ the cone was completed with extreme slowness, its vertex lying on the axis about 1 in. from R_1 . Thus, in this example, with the rotational speed not far above the minimum at which

régime III could exist, the downstream velocity in the central part of the cross-section was feeble, but even when the speed was raised to 40 rev/min, the length of downstream travel at the axis was not noticeably increased. With injection at R_3 , the motion at values of r/a down to 0.5 was downstream with a slight swirl that was soon lost; at $r/a = 0.5$ some of the colour, after going an inch, returned at a slightly smaller radius and reached the nearer joint, then being at a larger radius. At $r/a = 0.4$ and less, the motion was upstream just into the drum at a radius greater than that of injection; evidently the flow was passing outside the cone mentioned above.

The most convenient way of exploring the blocked half of the tubing was to continue to inject into the right-hand tube, the bung being transferred to the outlet of that tube, but the observations will be recorded as if they had been made in the left-hand portion. At L_1 and $r/a = 0.9$, the colour vigorously swirled away from the drum, and after reaching L_2 it returned at a smaller radius and passed through the drum, losing most of its swirl as it did so. At $r/a = 0.7$ the flow was directly into the drum as far as row -5 where it was caught in the stream nearer the wall and turned back into the blocked tube; eventually a trace returned into the drum at a smaller radius. At $r/a = 0.5-0.3$ all the colour entered the drum; some got through to the right-hand tube at a larger radius with a velocity that was low compared with the brisk motion of the rest which was returned towards L_1 . At $r/a = 0.2$ and less, the whole of the colour passed through the drum and emerged in two streams, one moving slowly at a large radius and the other with almost zero swirl and axial velocities close to the axis. This striking effect was particularly well marked at $r/a = 0$. The colour all reached the central plane where most moved quite abruptly to a larger radius and entered the right-hand tube. After it had gone, the remainder could be seen near the axis; it advanced with extreme slowness and finally came to rest about 1 in. downstream from R_1 , which was the position of the apex of the cone described in the previous paragraph. All the examples of régime III that were closely examined showed this division of the stream into two very different parts. At L_3 it was only extremely close to the wall that any flow existed away from the drum; the stream swirled towards the bung for $\frac{1}{2}$ in. and then crossed the section at a much smaller radius. At $r/a = 0.9-0.4$ the colour moved initially without swirl towards both the axis and the drum, but later in its travel it picked up swirl and most of it entered the drum. At still smaller radii the colour moved towards the drum with swirl that was evidently derived from the flow near the wall. An attempt to inject colour externally was unsuccessful because the inward velocity was so slow that the colour diffused before it entered the drum.

To discover to what extent the position of the blockage altered the régime

at R_1 , a sliding bung was placed in the left-hand tube, and two sets of experiments were made at constant V and R and varying positions of the blockage. In the first, at $V = 1,420$ and $R = 92$, régime III was found when the blockage was 4, $8\frac{1}{2}$, and 16 in. from the nearer joint, and II when this distance was 0, 1, 2 in.; in a trial at 3 in. the axial velocity near the axis was imperceptible. In the second, at $V = 2,360$, $R = 84$, the limit between the two régimes was at 4 in. Finally, the appearance of régime II when the blockage was near the drum was confirmed by fixing the bung as close to the joint as possible and making observations at constant $R = 116$ and varying rotational speed. Over the range 470–1,890 of V that was examined, only régime II occurred. At the highest speed the motion within the drum at small radii was very slow, and colour injected at the axis, after approaching the bung, moved into a clearly marked zone in the form of a conical frustum with its smaller end close to the bung and its larger near row +5. From the larger end the colour was whirled away downstream.

5. Experiments on symmetrical flow with a perforated bung in both the fixed tubes (arrangement C)

A sliding bung with a central $\frac{1}{4}$ -in. hole was now placed in both fixed tubes at a distance 8 in. from the nearer joint, thus the apparatus was converted into a primitive kind of double-ended swirl chamber. The results are plotted in Fig. 4. Along the axes the observations at R_1 were the same as before, and at a very low value of R there was, with rising V , the same transformation from régime I to II only, that is shown in Fig. 2. Under these conditions it seems that the bungs were too far away to have much influence on the flow at R_1 . At greater values of R , however, régime III was visible, the necessary rotational speed for its appearance becoming less as R was increased. At first, at $R = 60$ approximately, régime III was of the feeble kind described in section 4, in which the downstream travel over the central zone was small. But above $R = 200$, in all the examples of régime III that were observed, the central flow consisted of a vigorous and clearly defined stream that passed through the hole in the bung.

At $V = 1,990$, $R = 260$, with external injection at row 0, the colour divided symmetrically and soon formed into a diffuse cylinder at half radius extending over the length of the drum; the cylinder faded as colour was lost downstream from its ends, but none moved inwards to supply the discharge along the axis. At rows +2 and +5 the colour was rapidly whirled away downstream at a large radius. With injection at R_1 , the colour at the largest radii moved downstream and was quickly diffused, but as r/a was decreased below 0.7, reversed flow made its appearance. At $r/a = 0.5$ all the colour started upstream; some of it was removed downstream at a larger

radius, but the rest entered the drum in a rather diffused state and later formed itself into a cylinder of small radius extending from row +5 to R_1 and persisting for several seconds. At $r/a = 0.3$ the initial axial velocity was slight, nevertheless the cylinder formed again, extending quickly downstream to the hole and slowly up into the drum in a more diffuse state. At

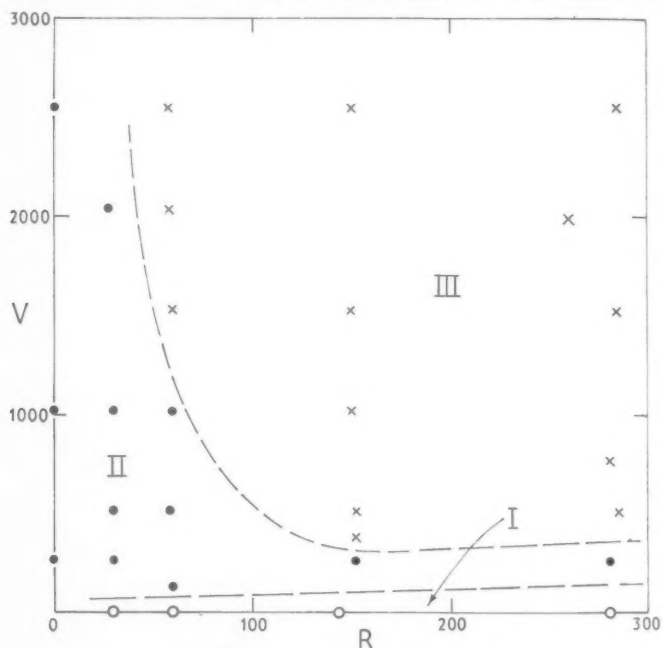


FIG. 4. Régime diagram for arrangement C

- Régime I
 ● " II
 × " III

smaller radii the motion was entirely downstream in the form of a sharply defined helix of small pitch that disappeared through the hole. Occasionally the filament bounced back out of the hole at a slightly greater radius, indicating that reverse flow existed there also.

6. Experiments on asymmetrical flow produced by blocking the left-hand tube close to the drum and inserting a perforated bung in the right-hand tube (arrangement D)

Finally a single-ended swirl chamber was simulated by placing in the left-hand tube a sliding unperforated bung as close as possible to the drum and in the right-hand tube a sliding bung with a $\frac{1}{4}$ -in. hole at a distance of

8 in. from the drum. The results are plotted in Fig. 5, which shows that again régime III was readily obtained. As in Fig. 4, for a rather small value of R régimes I, II, and III appeared in turn as the speed was increased, the onset of the change from II to III being again indicated by the decay and

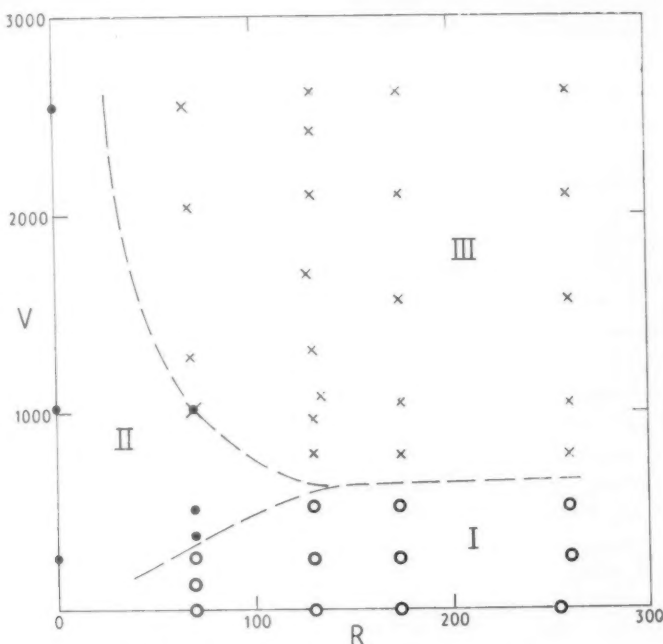


FIG. 5. Régime diagram for arrangement D

- Régime I
● " II
× " III

eventually the reversal of the upstream velocity at the axis. At higher values of R the intermediate régime II was not observed.

The details follow of what was seen of régime III at $V = 1,080$ (16 rev/min), $R = 134$ ($W = 12.4$ in./min). With injection at R_1 the colour inserted at $r/a = 0.9$ went downstream at a high axial velocity and diffused near R_3 . At $r/a = 0.8$ after moving downstream a couple of inches, some of the colour came back at a smaller radius and entered the drum. The same was seen at $r/a = 0.7$ except that a portion of the band that reached the drum was carried briskly downstream at a larger radius. Between $r/a = 0.6$ and 0.4 all moved towards the drum and some remained stagnant inside it; the remainder was removed downstream at a larger radius when

$r/a = 0.6$ and at a smaller radius when $r/a = 0.4$. The motion at $r/a = 0.3$ was in both directions, part going very slowly to the hole and the rest far into the drum at an increasing radius. For smaller values of r/a the colour moved downstream with its axial velocity increasing as it approached the hole.

With external injection at row +5 the colour swirled rapidly downstream close to the wall and diffused beyond R_3 . At row 0 the downstream axial velocity was at first so slight that the helix of colour that formed was barely distinguishable from a cylinder. At row -5 the colour moved into contact with the face of the bung where it could be seen in a diffuse state. Then, after a pause of several seconds, it gathered itself into three sharply defined streams 120° apart at a radius of about $\frac{2}{3}$ in., which passed with considerable angular velocity through the drum and far down the right-hand tube. A horizontal photograph, taken with a vertical camera and a submerged mirror inclined at 45° , is reproduced in Fig. 6, which shows the three streams with their heads almost in line when their development was well under way. The effect was found to be independent within wide limits of the duration and intensity of the injection, but one of the streams was faint when the duration was appreciably less than the time of one revolution of the drum; no alteration in the effect was noticed when simultaneous injection was made from a second probe fixed over the same row and distant about seven holes (42°) from the first probe. Similar results were observed when injections were made at rows -4, -3, and -2. The number of streams seen in many repetitions of the experiment was three, but four occurred on a few occasions when, apparently, conditions had not become steady after alterations in speed and discharge.

To elucidate this matter further, the discharge was left unchanged and observations were made at various speeds with injection at row -5. At 8 rev/min (régime I) the colour again collected up against the bung, and after a pause a helix of increasing diameter emerged giving rise to a cone of colour that passed downstream, though traces long remained visible in the drum. Thus the single point of injection gave rise to a single stream. The same occurred at 12 rev/min where régime III had just set in. At 24 rev/min the three streams were not quite as sharply defined as at 16 rev/min; at 32 rev/min they were definitely more diffuse, and whether they numbered three or four could not be decided. When this set of experiments was repeated with R increased to 216, it was found at the lower speeds that the cone tended to concentrate into streams and, as it passed through the drum, it threw out streams ahead. However, at 32 rev/min three sharply defined streams emerged right from the start, as in Fig. 6.

In conclusion, the perforated bung was removed, and thus the kind of

blocka
same c
extern
seen at
from t
both i
and th

7. Fu

Figs
condit
blocka
was ac
the ve
in a n
tube is
but th
for by
the wa
lateral
directi
seen in
stimul
that o
becaus
it cann
the fac
that in
some c
downs
the so
in C
annula
forate
tube i
radial
The
For th
first t
stream
in part

blockage was restored that is considered at the end of section 4. With the same discharge and swirls that are mentioned there, tests were made with external injection at row -5. Under these conditions régime II had been seen at R_1 , and now inside the drum separate streams of colour again sprang from the neighbourhood of the bung. But they were much more irregular both in spacing and in length than when the perforated bung was in place, and their number varied from two to four.

7. Further discussion of results

Figs. 2-5 differ widely, except along the axes, although the boundary conditions at entry remained unaltered; thus the arrangement of the blockage was of paramount importance. The chief object of the experiments was achieved in that they showed that régime III could be produced, but the velocities that were reached were small compared with those that occur in a nozzle with an air core. It appears that viscous swirling motion in a tube is so complicated that little can be done to explain it in general terms, but the existence of régime II at zero or small discharge can be accounted for by the centrifugal action inside the drum which caused the pressure near the wall to exceed that at the axis. Thus the water near the wall was forced laterally into the fixed tubing, and at the axis a current in the reverse direction was set up. Régime III is harder to explain. In the vigorous form seen in arrangements *C* and *D* the existence of the holes in the bungs stimulated a strong downstream flow close to the axis. Now it is well known that on the back wall of a single-ended swirl chamber there is inward flow because the liquid is retarded near the wall, its swirl being reduced so that it cannot maintain its position in the centrifugal field of force. Hence over the faces of the perforated bungs an inward stream was present, and it seems that in the narrowing region between this and the stream close to the axis some of the water was forced to move upstream. The origin of the central downstream current varied with the arrangement of the blockage. In *B* the source was the blocked left-hand tube in which régime II existed, and in *C* and *D* the current seemed to arise from the inner surface of the reverse annular stream. In *A*, where régime III was not seen, there were no perforated bungs to enhance the downstream motion on the axis nor a blocked tube in which it might develop; and although it was looked for, inward radial flow from row 0 remained undetected.

The effect shown in Fig. 6 does not seem to have been previously noticed. For the reason set out above, the colour inserted outside row -5 moved at first to the face of the neighbouring bung, but its emergence in separate streams was unexpected. The passage of these streams past R_1 accounts in part for the unsteadiness of the colour bands injected there.

REFERENCES

1. A. M. BINNIE and J. D. TEARE, *Proc. Roy. Soc. A*, **235** (1956) 78.
2. J. B. NUTTALL, *Nature*, **172** (1953) 582.
3. L. TALBOT, *J. App. Mech.* **21** (1954) 1.

A JE

By

The
past a
when a
to the
For the
is repre
sheet.
An ite
and th
jet sha
centre
a spec

1. In

We s
infini
zero
on th
motio
vorte
disco
follow
fluid
turn,
jet re
the a
will
main
at in
the a

2. T

Su
lowe
the r
[Qu

A JET DEFLECTED FROM THE LOWER SURFACE OF AN AEROFOIL

By H. J. DAVIES and A. J. ROSS (*University of Southampton*)

[Received 26 July 1956]

SUMMARY

The two-dimensional, steady motion of an infinite, perfect, incompressible fluid past a thin symmetrical aerofoil at zero incidence to the mainstream is considered when a jet issues from a general point on the lower surface of the aerofoil at an angle to the chord. The motion in both the mainstream and jet is taken to be irrotational. For the determination of the mainstream flow about the aerofoil and jet, the aerofoil is represented by a single plane vortex sheet and the jet by a single curvilinear vortex sheet. The strength of the latter is derived by considering the flow within the jet. An iterative method of solution is suggested for the evaluation of the true jet shape and the forces on the aerofoil; the first step requires the assumption of an approximate jet shape. As an example, the lift, pressure distribution, moment coefficient, and centre of pressure are calculated for a jet exit position close to the trailing-edge with a specific angle of ejection and momentum coefficient.

1. Introduction

We shall consider the two-dimensional, irrotational, steady motion of an infinite, perfect, incompressible fluid past a thin, symmetrical aerofoil at zero incidence to the mainstream when a jet issues from a general point on the lower surface of the aerofoil at an angle to the chord. The fluid motion in the jet is assumed to be irrotational and bounded by curvilinear vortex sheets across which the pressure is continuous, and the velocity discontinuous. At infinity the jet will be parallel to the mainstream, as follows at once from momentum considerations. The condition that the fluid is perfect implies no mixing of the jet with the mainstream which, in turn, implies constant total pressure heads H and J in the mainstream and jet respectively. Under these circumstances, the total lift experienced by the aerofoil, i.e. the pressure lift plus the component of the jet reaction will be given by $\rho U \sum (\Gamma + \gamma)$ where ρ is the density of the fluid, U the mainstream velocity, and $\sum (\Gamma + \gamma)$ the total circulation around a contour at infinity (1, 2). Hence, it is required to find the sum of the vorticity on the aerofoil $\sum \Gamma$ and on the jet boundaries $\sum \gamma$.

2. The mathematical model

Suppose a jet of small cross-section $B_1 B_2$ issues at an angle τ_0 from the lower surface of a thin symmetrical aerofoil which is at zero incidence to the mainstream U as shown in Fig. 1. The aerofoil is replaced by a vortex

sheet of strength Γ per unit length measured clockwise, and the jet boundaries B_1C_1 , B_2C_2 by vortex sheets of strengths γ_1 and γ_2 per unit arc length, also measured clockwise.

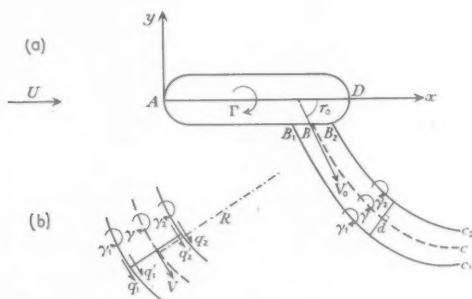


FIG. 1. (a) The physical plane, (b) jet section

Consider the one-dimensional flow across any section d of the jet across which the mean velocity is \bar{V} and the mean pressure p . Subscripts 0 and ∞ will be used to denote conditions of the jet at the exit and infinity. As B_1 and B_2 are stagnation points in the mainstream the pressure at the exit of the jet will be equal to the total head H of the mainstream. Hence

$$\frac{1}{2}\rho(\bar{V}^2 - \bar{V}_0^2) = H - p > 0, \quad \text{i.e. } \bar{V} > \bar{V}_0.$$

Thus, in the absence of mixing, the jet boundaries converge from the exit to infinity. The square of the ratio of the jet cross-sections d_0 and d_∞ , from continuity considerations, is given by $d_0^2/d_\infty^2 = \bar{V}_\infty^2/\bar{V}_0^2$, which, by application of Bernoulli's energy equation, may be written as $d_0^2/d_\infty^2 = 1 + U^2/\bar{V}_0^2$. Thus, it is seen that if the jet velocity \bar{V}_0 is very much greater than the mainstream velocity U , the jet may be regarded as having a constant cross-section. Finally, following Taylor (3) and Spence (4), the limit of the jet width is taken in such a way that the mass flow tends to zero but the momentum remains finite and constant. The jet is then represented by a single vortex sheet BC of strength γ per unit arc length, across which the pressure as well as the velocity is discontinuous.

3. The vorticity on the jet boundary

A section of the jet is shown in Fig. 1 (b), q and q' being the velocity outside and inside the jet respectively, p the pressure, and the subscripts 1 and 2 referring to the boundaries B_1C_1 and B_2C_2 respectively.

The strengths per unit arc length of the vortex sheets B_1C_1 and B_2C_2 regarded as boundaries of the flow in the mainstream will be $\gamma_1 = -q_1$, $\gamma_2 = q_2$. If corresponding increments of arc lengths of B_1C_1 , B_2C_2 , and

BC are taken to be equal the vortex density on BC will be given by $\gamma = \gamma_1 + \gamma_2 = q_2 - q_1$. As the pressure across B_1C_1 and B_2C_2 is continuous and the total heads of the jet and mainstream are constant an application of Bernoulli's energy equation across the section shown in Fig. 1 (b) gives

$$q_2^2 - q_1^2 = q_2'^2 - q_1'^2 = 2(p_1 - p_2)/\rho$$

so that

$$q_2' - q_1' = \frac{p_1 - p_2}{\rho \bar{V}}, \quad \text{where } \bar{V} = \frac{1}{2}(q_1' + q_2')$$

$$= \frac{d \operatorname{grad} p}{\rho \bar{V}}, \quad \text{since } d \text{ is assumed small,}$$

$$= \frac{d \bar{V}^2}{\bar{V} R}$$

by consideration of the acceleration normal to the jet, R being the radius of curvature of BC . Hence

$$\gamma = \frac{\bar{V}^2 d}{\bar{U} R}, \quad (3.1)$$

where $\bar{U} = \frac{1}{2}(q_1 + q_2)$ is taken to be constant and equal to the mainstream velocity U . It has already been stated that the model considered is one in which the momentum of the jet remains finite and constant. So that $\rho \bar{V}^2 d = \rho_0 \bar{V}_0^2 d_0$, where ρ_0 , the density of the jet fluid, is taken to be equal to ρ . By defining the non-dimensional jet momentum coefficient as $C_J = \rho_0 \bar{V}_0^2 d_0 / \frac{1}{2} \rho U^2 c$, where $c = AD$ the chord length, (3.1) can then be written as

$$\gamma = \frac{1}{2} U C_J c / R. \quad (3.2)$$

4. The vorticity on the aerofoil surface

Consider rectangular axes in which Ox is parallel to and in the direction of the mainstream and Oy is normal to the aerofoil chord AD , with origin at the leading edge A (see Fig. 1). One of the boundary conditions of the flow in the mainstream is that the velocity normal to the aerofoil surface is zero. From consideration of the model described in section 2 this condition gives

$$\frac{1}{2\pi} \int_0^c \frac{\Gamma(l) dl}{x-l} + \frac{1}{2\pi} \int_{BC} \frac{\gamma(x-X) ds}{(x-X)^2 + Y^2} = 0 \quad (0 \leq x \leq c), \quad (4.1)$$

where l is the distance of a general point on the chord from the leading edge and X, Y the coordinates of a general point on BC corresponding to an arc length s measured from B . The second integral in this equation may be interpreted as the downwash on the aerofoil due to the vortex sheets bounding the jet and will be denoted by $-(1/2\pi)\omega(x)$.

Consider the integral

$$\int_0^c \frac{x^{\frac{1}{2}}(c-x)^{\frac{1}{2}} dx}{(\lambda-x)(x-l)}$$

which, by substituting $x = c \sin^2 \theta$, becomes

$$2 \int_0^{\frac{1}{2}\pi} \frac{\sin^2 \theta \cos^2 \theta d\theta}{\{(\lambda/c) - \sin^2 \theta\} \{\sin^2 \theta - (l/c)\}} = \pi, \quad \text{if } 0 \leq \lambda \leq c. \quad (4.2)$$

Thus the operator

$$\int_0^c \frac{x^{\frac{1}{2}}(c-x)^{\frac{1}{2}}}{(\lambda-x)} dx$$

is seen to be a null transform of order one to the kernel $1/(x-l)$, in the interval $0 \leq x \leq c$. By the application of this transform to the equation

$$\int_0^c \frac{\Gamma(l)}{x-l} dl = \omega(x)$$

it is seen that

$$\int_0^c \Gamma(l) \int_0^c \frac{x^{\frac{1}{2}}(c-x)^{\frac{1}{2}}}{(\lambda-x)(x-l)} dx dl + R(\lambda) = \int_0^c \frac{\omega(x)x^{\frac{1}{2}}(c-x)^{\frac{1}{2}}}{\lambda-x} dx, \quad (4.3)$$

where $R(\lambda)$ is the residual, arising from the singularity in the integrand when the order of integration is reversed. Thus

$$R(\lambda) = \lim_{\epsilon \rightarrow 0} \left[- \int_{l=\lambda-\epsilon}^{\lambda+\epsilon} \Gamma(l) \int_{x=\lambda-\epsilon}^{\lambda+\epsilon} \frac{x^{\frac{1}{2}}(c-x)^{\frac{1}{2}}}{(\lambda-x)(x-l)} dx dl + \int_{x=\lambda-\epsilon}^{\lambda+\epsilon} \frac{x^{\frac{1}{2}}(c-x)^{\frac{1}{2}}}{\lambda-x} \int_{l=\lambda-\epsilon}^{\lambda+\epsilon} \frac{\Gamma(l)}{x-l} dl dx \right]$$

which, on substituting $l = \epsilon \xi + \lambda$, $x = \epsilon \eta + \lambda$ and proceeding to the limit, gives

$$R(\lambda) = -\pi^2 \Gamma(\lambda) \lambda^{\frac{1}{2}} (c-\lambda)^{\frac{1}{2}} \quad (4.4)$$

(see reference 5). Thus the inversion of the integral equation (4.1) is, from (4.2) to (4.4), given by

$$\Gamma(\lambda) = \frac{1}{\pi^2 \lambda^{\frac{1}{2}} (c-\lambda)^{\frac{1}{2}}} \left[\pi \int_0^c \Gamma(l) dl - \int_0^c \frac{\omega(x)x^{\frac{1}{2}}(c-x)^{\frac{1}{2}}}{\lambda-x} dx \right], \quad (4.5)$$

where

$$\omega(x) = - \int_{BC} \frac{\gamma(x-X)}{(x-X)^2 + Y^2} ds.$$

The constant $\int_0^c \Gamma(l) dl$ is determined from the condition that B is a stagnation point of the mainstream. The condition that the velocity normal to AD is zero has been satisfied above. Hence, to satisfy the condition that B is a stagnation point, it is sufficient to write down the condition that the x -component of the velocity at that point, taken to be $x = b$, is zero, i.e.

$$-U + \frac{\Gamma(b)}{2} + \frac{1}{2\pi} P \int_{BC} \frac{\gamma Y}{(b-X)^2 + Y^2} ds = 0, \quad (4.6)$$

where P denotes the Cauchy principal value of the integral. This gives a value for $\Gamma(b)$ which may be substituted in equation (4.5), writing $z = b$, in order to obtain the value of $\int_0^c \Gamma(l) dl$.

It is of interest to note that the significant difference between the above model and that adopted for treatment by thin-aerofoil theory (4) is the position of the jet exit, as this changes the boundary condition required for the determination of the constant arising in the solution of the integral equation for the vorticity on the aerofoil. This results in a different order of infinity of the vorticity on the aerofoil at the trailing edge, and an infinite difference in the vorticity on the jet mean line there, although the total jet vorticity is unchanged. However, it can be shown that the solution when the jet is taken at the trailing edge as in reference (4) is the limit as the jet position tends to the trailing edge in the solution contained in this paper.

5. The lift

The total circulation within the mainstream is given by

$$\int_0^c \Gamma(l) dl + \int_{BC} \gamma ds,$$

and so, as stated in section 1, the lift experienced by the aerofoil will be given by

$$L = \rho U \left[\int_0^c \Gamma(l) dl + \int_{BC} \gamma ds \right]. \quad (5.1)$$

By substituting for γ from (3.2) and transforming the variable of integration from arc length to tangent slope so that $\delta s = R \delta \tau$, the total vorticity on the jet boundaries may be written as

$$\int_{BC} \gamma ds = \frac{1}{2} C_J \tau_0 U c. \quad (5.2)$$

The method of calculating the circulation around the aerofoil has been given in section 4. By substituting for γ and δs , given above, in equations (4.5) and (4.6), together with $\lambda = b$ in (4.5), and eliminating $\Gamma(b)$ between these two equations, the aerofoil circulation is then given by the equation

$$\int_0^c \Gamma(l) dl = \pi b^{\frac{1}{2}}(c-b)^{\frac{1}{2}} \left[2U - \frac{C_J U c}{2\pi} \int_{\tau_0}^0 \frac{Y d\tau}{(b-X)^2 + Y^2} \right] + \frac{1}{\pi} \int_0^c \omega(x) \frac{x^{\frac{1}{2}}(c-x)^{\frac{1}{2}}}{b-x} dx, \quad (5.3)$$

where
$$\omega(x) = -\frac{1}{2} C_J U c \int_{\tau_0}^0 \frac{x-X}{(x-X)^2 + Y^2} d\tau. \quad (5.4)$$

It is seen from equation (5.3) that another boundary condition is required to determine the remaining unknown quantity, namely the equation of the limit line BC of the jet. The outstanding condition is that the velocity of the mainstream normal to BC is zero. By taking X^* , Y^* to be the co-ordinates of a general point and τ^* the slope at that point of the line BC given by $X = X(\tau)$, $Y = Y(\tau)$, this condition may be written in the form

$$\begin{aligned} \sin \tau^* \left[U + \frac{1}{2\pi} \left(\int_0^c \frac{\Gamma(l) Y^*}{(X^*-l)^2 + Y^{*2}} dl + \right. \right. \\ \left. \left. + \left(\frac{1}{2} C_J U c \right) P \int_{\tau_0}^0 \frac{Y^* - Y}{(X^* - X)^2 + (Y^* - Y)^2} d\tau \right) \right] \\ = \frac{\cos \tau^*}{2\pi} \left[\int_0^c \frac{\Gamma(l) (X^* - l)}{(X^* - l)^2 + Y^{*2}} dl + \right. \\ \left. + \left(\frac{1}{2} C_J U c \right) P \int_{\tau_0}^0 \frac{(X^* - X)}{(X^* - X)^2 + (Y^* - Y)^2} d\tau \right]. \quad (5.5) \end{aligned}$$

It does not seem possible to obtain X and Y as explicit functions of τ . However, if an approximation to the equation of BC can be obtained, then equations (5.1) to (5.3) will give a first solution for lift whilst equation (5.5) offers a means of iteration.

6. The limit line

As an approximation to the equation of the limit line BC of the jet it was decided to use the solution of the source type flow given by Woods (6).

On the assumption that jets of equal momentum and at equal angles of ejection have the same influence on the mainstream, the equation of the mean line of Woods's source type flow, having the same momentum and angle of ejection as that of the jet, was taken as an approximation to the equation of BC . The downwash $\omega(x)$ on the aerofoil may now be evaluated numerically and the circulation around the aerofoil obtained from equation (5.3).

In order to obtain a new position of the jet limit line the flow in the region of the assumed curve must be considered. Now $\tan \tau^*$, which may be derived from equation (5.5), will be the slope of the streamlines cutting the assumed curve and must, with each successive iteration, tend to the slope of the limit line itself. The velocity field of the flow indicates the true position of the jet but the computation involved is too lengthy. However, it is possible to derive a measure of the displacement along the normal of the assumed curve from the normal velocity. If n is the normal displacement, s the arc length, q_n the normal velocity, and q_t the tangential velocity at the point on the assumed curve then $dn/ds \simeq q_n/q_t$, i.e.

$$n \simeq \int \frac{q_n}{q_t} ds.$$

In this way a new equation to the limit line may be obtained, since if X_1, Y_1 are points on the assumed curve and X_2, Y_2 points on the corrected curve, then

$$X_2 = X_1 + n \sin \tau \quad \text{and} \quad Y_2 = Y_1 + n \cos \tau.$$

7. Moment coefficient and pressure distribution

The pressure force contribution to the moment coefficient about the leading edge of the aerofoil will be given by

$$\begin{aligned} C_M &= \frac{1}{\frac{1}{2}\rho U^2 c^2} \int_0^1 (p_l - p_u) x \, dx \\ &= \int_0^1 (q_u^2 - q_l^2) x \, dx \quad (\text{taking unit values of } U \text{ and } c), \end{aligned}$$

where subscripts l and u denote lower and upper surfaces and

$$q_u = 1 + U_\gamma + \frac{1}{2}\Gamma(x), \quad q_l = 1 + U_\gamma - \frac{1}{2}\Gamma(x), \quad U_\gamma = \frac{C_J}{4\pi} \int_1^0 \frac{Y^*}{(X^* - x)^2 + Y^{*2}} d\tau.$$

Thus,

$$C_M = \frac{1}{2} \int_0^1 (1 + U_\gamma) \Gamma(x) x \, dx.$$

The pressure coefficient is given by $C_p = 1 - q^2$, where q has the values q_∞ and q_l given above.

8. Numerical analysis

For the numerical work it is more convenient to use a dimensionless form of the equations; lengths are given as ratios of the aerofoil chord and velocities as ratios of the mainstream velocity at infinity. The equations of the preceding paragraphs will remain unaltered in form if c and U are replaced by unity.

The evaluation of the principal value of integrals involved in this work is mainly standard work. However, there occur two integrals the evaluation of which calls for an explanation. Consider the integrals

$$\int_{\tau_0}^0 \frac{X^* - X}{(X^* - X)^2 + (Y^* - Y)^2} d\tau \quad (8.1)$$

and

$$\int_{\tau_0}^0 \frac{Y^* - Y}{(X^* - X)^2 + (Y^* - Y)^2} d\tau \quad (8.2)$$

occurring in equation (5.5). Let $P^*(X^*, Y^*, \tau^*)$ be any point of the curve BC , $P(X, Y, \tau)$ a neighbouring point of arc length s and chord length h from P^* and θ the angle between P^*P and the x -axis (see Fig. 2).

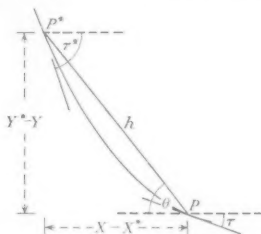


FIG. 2. Jet mean line geometry

The integrands of (8.1) and (8.2) may be written as $(\cos \theta)/h$ and $-(\sin \theta)/h$, respectively, where in the neighbourhood of P^* , $h \simeq s \simeq -R^*(\tau^* - \tau)$. Thus the orders of infinity of the integrands of (8.1) and (8.2) are

$$-\frac{\cos \tau^*}{R^*(\tau^* - \tau)} \quad \text{and} \quad \frac{\sin \tau^*}{R^*(\tau^* - \tau)},$$

and these functions are used to 'subtract out' the infinity of the integrands. Hence, the value of the integrand of (8.1) at $\tau = \tau^*$ is given by

$$\lim_{\tau \rightarrow \tau^*} \left[\frac{X^* - X}{(X^* - X)^2 + (Y^* - Y)^2} + \frac{\cos \tau^*}{R^*(\tau^* - \tau)} \right].$$

By writing

$$\tau = \tau^* + \epsilon,$$

so that

$$X = X^* + \epsilon X_\tau + \frac{1}{2} \epsilon^2 X_{\tau\tau} + O(\epsilon^3)$$

and

$$Y = Y^* + \epsilon Y_\tau + \frac{1}{2} \epsilon^2 Y_{\tau\tau} + O(\epsilon^3),$$

where

$$X_\tau = -R \cos \tau, \quad Y_\tau = R \sin \tau,$$

$$X_{\tau\tau} = -R_\tau \cos \tau + R \sin \tau, \quad Y_{\tau\tau} = R_\tau \sin \tau + R \cos \tau,$$

the limit becomes equal to

$$\left[-\frac{\sin \tau^*}{2R} - \frac{\cos \tau}{2R^2} \frac{dR}{d\tau} \right]_{\tau=\tau^*}.$$

The integral (8.1) may then be written in the form

$$(8.1) \quad \int_{\tau_0}^0 \left(\frac{X^* - X}{(X^* - X)^2 + (Y^* - Y)^2} + \frac{\cos \tau^*}{R^*(\tau^* - \tau)} \right) d\tau - \frac{\cos \tau^*}{R^*} \log \left| \frac{\tau_0 - \tau^*}{\tau^*} \right|.$$

Similarly, (8.2) may be written in the form

$$(8.2) \quad \int_{\tau_0}^0 \left(\frac{Y^* - Y}{(X^* - X)^2 + (Y^* - Y)^2} - \frac{\sin \tau^*}{R^*(\tau^* - \tau)} \right) d\tau + \frac{\sin \tau^*}{R^*} \log \left| \frac{\tau_0 - \tau^*}{\tau^*} \right|.$$

9. Numerical results

In the numerical example it was decided to take the jet near to the trailing edge in order that the result obtained could be compared with experimental and theoretical results (4) already available at the time.

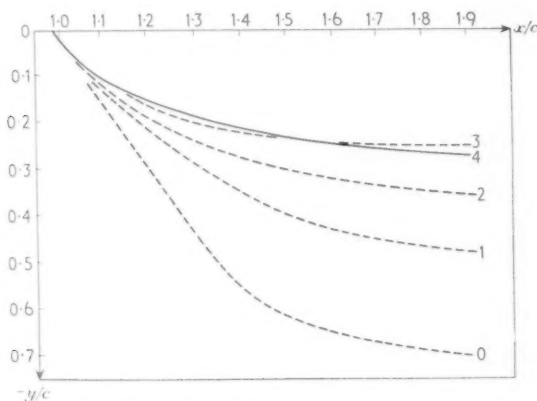


FIG. 3. Mean line iteration ($\tau_0 = 55.5^\circ$, $C_f = 0.5$)

The values $\tau_0 = 55.5^\circ$ and $b/c = 0.991$ were therefore taken in our calculations to correspond to an early model which had then been tested at the

N.G.T.E.; these experiments have now been superseded by more accurate tests on a closely similar model with $\tau_0 = 58.1^\circ$ (7).

Theoretical values were obtained for the lift, pressure distribution, and 'pressure-force' contribution to the moment corresponding to a jet coefficient $C_J = 0.5$. The initial curve, marked 0 in Fig. 3, for the limit line of the jet as given by the corresponding source type flow taken from reference (6) was found to be a poor approximation to the true mean line, and the iteration was repeated four times, marked 1, 2, 3, 4 in Fig. 3, before satisfactory agreement between successive lift coefficients was obtained. The final theoretical value of the lift coefficient was 2.60 for $\tau_0 = 55.5^\circ$ as compared with an experimental value (7) of 2.9 for $\tau_0 = 58.1^\circ$, and with the value 2.75 for $\tau_0 = 60^\circ$ obtained by the theory of reference (4). The pressure distribution around a $12\frac{1}{2}$ per cent. thick aerofoil was calculated using the usual transformations from flat plate to circle to ellipse, and is shown in Fig. 4. The pressure-force contribution to the moment coefficient about the leading edge was calculated to be 1.42 and the centre of total lift at the point $x = 0.55$.

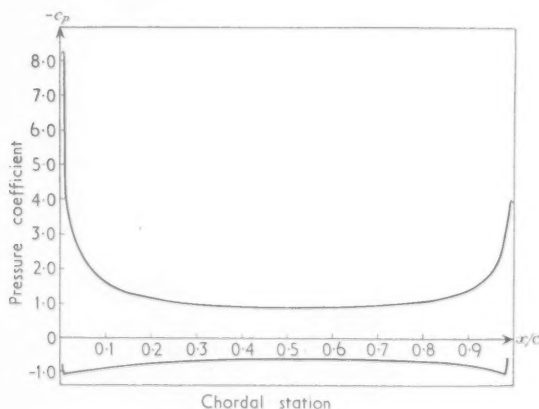


FIG. 4. Pressure distribution on a $12\frac{1}{2}\%$ thick ellipse ($\tau_0 = 55.5^\circ$, $C_J = 0.5$)

10. Conclusions

An iterative method of solution for the equations of the limit line of the jet and consequently for the pressure distribution, lift, moment coefficient, and centre of total lift, is given when the jet is at a general position on the lower surface of an aerofoil. The length of computation involved makes the method too cumbersome for general use unless the initial curve chosen to represent the limit line of the jet is very much nearer the true position

JET DE

than th
line of
be to c
commo
could t
experim

1. G. I.
2. J. H.
3. G. I.
Pu
4. D. A.
5. H. L.
6. L. C.
7. N. A.

than that taken in the example described in sections 6-7, namely, the mean line of the corresponding source type flow. A possible alternative would be to obtain the limit lines of the jet experimentally for a range of τ_0 for one common value of C_J . For a given τ_0 limit lines of the jet for successive C_J 's could then be obtained quite easily and quickly by iteration from the experimentally obtained limit lines.

REFERENCES

1. G. I. TAYLOR, Appendix to R and M 989, 1924.
2. J. H. PRESTON, A.R.C. R.M. 2957, 1956.
3. G. I. TAYLOR, *Offerts à Riabouchinsky. Memoires sur la Mecanique des Fluides.* Pub. Sci. et Tech. du Min. de l'Air, 1954, p. 313.
4. D. A. SPENCE, R.A.E. Rep. Aero 2568, 1955.
5. H. LOWMAN, M. A. HEASLET, and F. B. FULLER, N.A.C.A. Tech. Rep. 1054, 1951.
6. L. C. WOODS, *Quart. J. Mech. App. Math.* **9** (1956) 441.
7. N. A. DIMMOCK, N.G.T.E. Rep. R 175, 1955.

LAMINAR BOUNDARY LAYERS AT THE INTERFACE OF CO-CURRENT PARALLEL STREAMS

By O. E. POTTER

(*Department of Chemical Engineering, Manchester College of Science and
Technology and Manchester University*)

[Received 1 January 1956; revise received 4 October 1956]

SUMMARY

The approximate solution of Keulegan (1) for the steady flow of a stream of viscous incompressible fluid over another at rest is extended to the case where both fluids are moving co-current but at different velocities. This solution utilizes a sextic polynomial for the velocity distribution in the boundary layers. The solutions depend only on the ratio U_2/U_1 of the velocities of the two streams and on the product of the corresponding viscosity and density ratios. Numerical results are given for seven values of $\mu_2 \rho_2 / (\mu_1 \rho_1)$ at one value of U_2/U_1 . Lock (2) has published an exact solution with a numerical result for $\mu_2 \rho_2 / (\mu_1 \rho_1) = 1$, $U_2/U_1 = 0.501$ and the sextic polynomial solution is evaluated for this case also. Comparison indicates that in general the sextic polynomial is more accurate than the quartic polynomial but that the advantage is not great.

1. Introduction

DISTILLATION, gas absorption, and liquid-liquid extraction, three important chemical engineering operations, are all instances of mass transfer from one fluid phase to another, both phases being in motion. Little is known of the fundamental hydrodynamics of fluid-fluid interaction even where stability of the interface is assumed. Keulegan (1), utilizing a sextic polynomial for the velocity distribution in the laminar boundary layers, gave a solution applicable only in the case where the lower stream was at rest. Later Lock (2) obtained an exact solution applicable to the two cases—lower stream at rest, lower stream in motion. Lock also gave the momentum integral solutions for various assumed forms of the velocity distribution, including a quartic polynomial form which is the same as that used by Keulegan for his first approximation.

These various solutions to the laminar boundary layers give rise to corresponding solutions for mass transfer between the two fluid phases. Solutions to the mass transfer case have been obtained and will be published later.

Unfortunately such solutions can be only of limited value since one of the prime factors in fluid-fluid interaction and in the behaviour of fluids with a free surface is the question of the stability of the interface or free

surface. Various aspects of this question have been attacked by Lessen (3) and Yih (4), whose results confirm the expectation that instability sets in at rather low values of Reynolds numbers.

2. The boundary layer equations

Consider two incompressible fluids flowing parallel and co-current; let the upper stream be of density ρ_1 , viscosity μ_1 , and have velocity U_1 , and the lower stream have density ρ_2 , viscosity μ_2 , and velocity U_2 .

The origin is taken to be the point at which the two fluids first come in contact, the axis of x is horizontal and in the direction of motion and the axis of y vertically upwards (see Fig. 1). The corresponding components

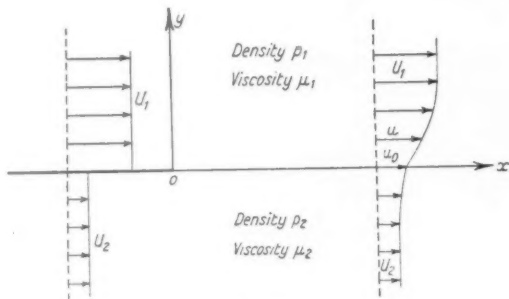


FIG. 1

of velocity are denoted by u and v . The boundary layer equations are then:

$$\text{upper fluid} \quad u \frac{\partial u}{\partial x} + v \frac{\partial u}{\partial y} = \nu_1 \frac{\partial^2 u}{\partial y^2}, \quad (1)$$

$$\text{lower fluid} \quad u \frac{\partial u}{\partial x} + v \frac{\partial u}{\partial y} = \nu_2 \frac{\partial^2 u}{\partial y^2}, \quad (2)$$

where ν_1 and ν_2 are the kinematic viscosities of the two fluids.

The equation of continuity is

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0 \quad (3)$$

so that there exists a stream function ψ such that

$$u = \frac{\partial \psi}{\partial y}, \quad v = -\frac{\partial \psi}{\partial x}. \quad (4)$$

Introducing the dimensionless variables

$$\eta_1 = \left(\frac{U_1}{\nu_1 x} \right)^{\frac{1}{2}} y, \quad \eta_2 = \left(\frac{U_1}{\nu_2 x} \right)^{\frac{1}{2}} y, \quad (5)$$

a solution is sought for in which:

$$\text{upper fluid} \quad \psi = (\nu_1 U_1 x)^{\frac{1}{2}} f_1(\eta_1), \quad (6)$$

$$\text{lower fluid} \quad \psi = (\nu_2 U_1 x)^{\frac{1}{2}} f_2(\eta_2). \quad (7)$$

Then we have

$$\text{upper fluid} \quad u = U_1 f_1'(\eta_1), \quad (8)$$

$$\text{lower fluid} \quad u = U_1 f_2'(\eta_2), \quad (9)$$

where accents denote derivatives:

$$\text{upper fluid} \quad v = \frac{1}{2} \left(\frac{U_1 \nu_1}{x} \right)^{\frac{1}{2}} \{ \eta_1 f_1'(\eta_1) - f_1(\eta_1) \}, \quad (10)$$

$$\text{lower fluid} \quad v = \frac{1}{2} \left(\frac{U_1 \nu_2}{x} \right)^{\frac{1}{2}} \{ \eta_2 f_2'(\eta_2) - f_2(\eta_2) \}, \quad (11)$$

$$\text{upper fluid} \quad \frac{\partial u}{\partial y} = U_1 \left(\frac{U_1}{\nu_1 x} \right)^{\frac{1}{2}} f_1''(\eta_1), \quad (12)$$

$$\text{lower fluid} \quad \frac{\partial u}{\partial y} = U_1 \left(\frac{U_1}{\nu_2 x} \right)^{\frac{1}{2}} f_2''(\eta_2). \quad (13)$$

Equations (1) and (2) then reduce to

$$2 \frac{d^3 f_1}{d\eta_1^3} + f_1 \frac{d^2 f_1}{d\eta_1^2} = 0 \quad (\text{upper fluid}) \quad (14)$$

$$\text{and} \quad 2 \frac{d^3 f_2}{d\eta_2^3} + f_2 \frac{d^2 f_2}{d\eta_2^2} = 0 \quad (\text{lower fluid}). \quad (15)$$

2.1. Boundary conditions

The boundary conditions (discussed in some detail by Lock (2)) which completely determine the solution of equations (14) and (15) are

$$u \rightarrow U_1 \quad \text{as } \eta_1 \rightarrow \infty, \quad \text{i.e. } f_1'(\infty) = 1, \quad (16)$$

$$u \rightarrow U_2 \quad \text{as } \eta_2 \rightarrow -\infty, \quad \text{i.e. } f_2'(-\infty) = U_2/U_1 = \lambda; \quad (17)$$

since the motion is steady, we have

$$f_1(0) = f_2(0) = 0. \quad (18)$$

For continuity of velocity,

$$f_1'(0) = f_2'(0), \quad (19)$$

and for continuity of tangential stress, we have

$$\rho_1 \nu_1^{\frac{1}{2}} f_1''(0) = \rho_2 \nu_2^{\frac{1}{2}} f_2''(0). \quad (20)$$

2.2. Exact solution

It is necessary to know the asymptotic forms which the solutions of the equation

$$2f'''(\eta) + f(\eta)f''(\eta) = 0$$

take when η is large. For the case $\lambda > 0$, Lock (2) wrote

$$\lim_{\eta \rightarrow -\infty} (f - \lambda \eta) = E,$$

where E is a constant, and for η large and negative obtained

$$f \sim \lambda \eta + E + \frac{k e^{-z^2}}{z^2} \left(1 - \frac{3}{2z^2} + \frac{15}{4z^4} - \dots \right) - \frac{1}{8} \frac{k^2 e^{-2z^2}}{\lambda^{\frac{1}{2}} z^5} + \dots, \quad (21)$$

$$f' \sim \lambda + 2k\lambda^{\frac{1}{2}} \operatorname{erf} z - \frac{1}{4} \frac{k^2 e^{-2z^2}}{z^4} + \dots, \quad (22)$$

$$f'' \sim k\lambda e^{-z^2} - \frac{1}{2} k^2 \lambda^{\frac{1}{2}} \frac{e^{-2z^2}}{z^3} + \dots \quad (23)$$

where k is a constant and

$$z = -\frac{1}{2} \lambda^{\frac{1}{2}} \left(\eta + \frac{E}{\lambda} \right) \quad (24)$$

with corresponding expansions for $f(\eta)$ when η is large and positive.

Lock then obtained by numerical integration a solution for $\lambda = 0.501$ and $\mu_2 \rho_2 / (\mu_1 \rho_1) = 1$. He also obtained solutions for $\lambda = 0$ (lower liquid at rest) at various values of $\mu_2 \rho_2 / (\mu_1 \rho_1)$, namely 5.965×10^4 , 100, 10, 1.

3. Approximate solution by the momentum equation

If the equations of motion (1) and (2) are integrated with respect to y between the limits 0 and δ_1 , 0 and $-\delta_2$ respectively, where δ_1 and δ_2 are the boundary layer thicknesses on either side of the interface, then the momentum equations are obtained:

$$\text{upper fluid} \quad \frac{\tau_0}{\rho_1} = \nu_1 \left(\frac{\partial u}{\partial y} \right)_{y \rightarrow 0+} = \frac{\partial}{\partial x} \int_0^{\delta_1} (U_1 - u) u \, dy, \quad (25)$$

$$\text{lower fluid} \quad \frac{\tau_0}{\rho_2} = \nu_2 \left(\frac{\partial u}{\partial y} \right)_{y \rightarrow 0-} = \frac{\partial}{\partial x} \int_0^{-\delta_2} (U_2 - u) u \, dy, \quad (26)$$

where τ_0 is the skin friction. In the upper fluid we write

$$\frac{u}{U_1} = \phi_1(\eta_1^*),$$

$$\eta_1^* = y/\delta_1.$$

Then equation (25) becomes

$$\frac{\nu_1}{U_1} \phi_1'(0) = \delta_1 \frac{d\delta_1}{dx} \int_0^1 (\phi_1 - \phi_1^2) d\eta_1^*,$$

whence

$$\frac{2\nu_1 x}{U_1} \phi_1'(0) = \delta_1^2 \int_0^1 (\phi_1 - \phi_1^2) d\eta_1^*. \quad (27)$$

For the lower fluid

$$\frac{u}{U_1} = \phi_2(\eta_2^*),$$

where

$$\eta_2^* = -y/\delta_2,$$

and equation (26) becomes

$$\frac{2\nu_2 x}{U_1} \phi_2'(0) = \delta_2^2 \int_0^1 (\lambda \phi_2 - \phi_2^2) d\eta_2^*. \quad (28)$$

3.1. Sextic and quartic polynomial velocity distributions

The velocity distribution used by Keulegan (1) was a sextic polynomial which is again used here:

upper fluid
$$\phi_1(\eta_1^*) = a_0 + \frac{1-a_0}{A_1} g_1(\eta_1^*), \quad (29)$$

where, if the interfacial velocity is u_0 , $a_0 = u_0/U_1$, and also A_1 is a constant, and

$$g_1(\eta_1^*) = (1+m_1)\eta_1^* - (1-\frac{16}{3}l_1 + 10m_1)\eta_1^{*3} + (\frac{1}{2} - 8l_1 + \frac{55}{2}m_1)\eta_1^{*4} + (\frac{16}{3}l_1 - 27m_1)\eta_1^{*5} + 9m_1\eta_1^{*6}, \quad (30)$$

where l_1 and m_1 are constants.

Similarly for the lower fluid

$$\phi_2(\eta_2^*) = a_0 + \frac{\lambda - a_0}{A_2} g_2(\eta_2^*), \quad (31)$$

where A_2 is a constant and

$$g_2(\eta_2^*) = (1+m_2)\eta_2^* - (1-\frac{16}{3}l_2 + 10m_2)\eta_2^{*3} + (\frac{1}{2} - 8l_2 + \frac{55}{2}m_2)\eta_2^{*4} + (\frac{16}{3}l_2 - 27m_2)\eta_2^{*5} + 9m_2\eta_2^{*6}, \quad (32)$$

where l_2 and m_2 are constants.

If the constants l and m are put equal to zero then (as will be seen later) $A_1 = \frac{1}{2} = A_2$ and equations (29) and (31) become

$$\phi_1(\eta_1^*) = a_0 + (1-a_0)(2\eta_1^* - 2\eta_1^{*3} + \eta_1^{*4}), \quad (33)$$

$$\phi_2(\eta_2^*) = a_0 + (\lambda - a_0)(2\eta_2^* - 2\eta_2^{*3} + \eta_2^{*4}). \quad (34)$$

3.2. Boundary conditions

The velocity distributions must satisfy the boundary conditions

$$\phi_1(1) = 1, \quad \phi_1'(1) = \phi_1''(1) = \phi_1'''(1) = 0,$$

$$\phi_2(1) = \lambda, \quad \phi_2'(1) = \phi_2''(1) = \phi_2'''(1) = 0,$$

$$\phi_1(0) = \phi_2(0) = a_0,$$

$$\phi_1''(0) = \phi_2''(0) = 0,$$

$$\frac{\mu_1}{\delta_1} \phi_1'(0) = -\frac{\mu_2}{\delta_2} \phi_2'(0).$$

Also, for $\eta_1^* = \eta_2^* = 0$ these differential equations must be satisfied when differentiated with respect to y .

3.3. Application of boundary conditions

Since $\phi_1(1) = 1$, then from equations (29) and (30)

$$(28) \quad 1 = a_0 + \frac{1-a_0}{A_1} \left(\frac{1}{2} + \frac{8}{15}l_1 + \frac{1}{2}m_1 \right). \quad (35)$$

$$\text{Hence,} \quad A_1 = \frac{1}{2} \left(1 + \frac{16}{15}l_1 + m_1 \right). \quad (36)$$

Similarly, for the lower stream, since $\phi_2(1) = \lambda$,

$$(29) \quad \lambda = a_0 + \frac{\lambda - a_0}{A_2} \left(\frac{1}{2} + \frac{8}{15}l_2 + \frac{1}{2}m_2 \right). \quad (37)$$

$$\text{Hence,} \quad A_2 = \frac{1}{2} \left(1 + \frac{16}{15}l_2 + m_2 \right). \quad (38)$$

$$\text{Then, since} \quad \phi_1'''(1) = 0 = \phi_2'''(1),$$

$$\text{it follows that} \quad 6 + 32l_1 + 60m_1 = 0 \quad (39)$$

$$\text{and} \quad 6 + 32l_2 + 60m_2 = 0. \quad (40)$$

Using the condition for the continuity of stress, viz.

$$\frac{\mu_1}{\delta_1} \phi_1'(0) = -\frac{\mu_2}{\delta_2} \phi_2'(0),$$

$$(31) \quad \text{we have} \quad \frac{\mu_1(1+m_1)(1-a_0)}{\delta_1 A_1} = -\frac{\mu_2(1+m_2)(\lambda-a_0)}{\delta_2 A_2}. \quad (41)$$

If the equation of motion (1) is differentiated with respect to y and the conditions at the interface $\eta_1^* = \eta_2^* = 0$ inserted, then

$$a_0 U_1 \frac{1-a_0}{A_1} (1+m_1) \left(-\delta_1 \frac{d\delta_1}{dx} \right) = \nu_1 \frac{1-a_0}{A_1} (-6 + 32l_1 - 60m_1)$$

which, on integration, yields

$$(33) \quad -a_0 \frac{1-a_0}{A_1} (1+m_1) = \frac{2\nu_1 x(1-a_0)}{\delta_1^2 A_1 U_1} (-6 + 32l_1 - 60m_1). \quad (34)$$

Using equation (39) this equation simplifies to

$$64l_1 = -a_0(1+m_1)\delta_1^2(U_1/2\nu_1 x). \quad (42)$$

Similarly, for the lower stream, we have

$$64l_2 = -a_0(1+m_2)\delta_2^2(U_1/2\nu_2 x). \quad (43)$$

3.4. Integration of the momentum equations

The sextic polynomial velocity distributions are used to integrate the momentum equations (27) and (28). For the upper stream

$$\frac{2\nu_1 x(1-a_0)}{\delta_1^2 U_1 A_1} = \frac{a_0(1-a_0)B_1}{A_1} + C_1 \left(\frac{1-a_0}{A_1} \right)^2, \quad (44)$$

where

$$B_1(1+m_1) = 0.15 + 0.26667l_1 + 0.214285m_1 \quad (45)$$

and

$$C_1(1+m_1) = 0.029365 + 0.0762l_1 + 0.07626m_1 + 0.0309l_1^2 + 0.07619l_1m_1 + 0.0401m_1^2. \quad (46)$$

Similarly, for the lower stream,

$$\frac{2\nu_2 x(\lambda - a_0)}{\delta_2^2 U_1 A_1} = B_2 \frac{\lambda - a_0}{A_2} a_0 + C_2 \left(\frac{\lambda - a_0}{A_2} \right)^2, \quad (47)$$

where

$$B_2(1+m_2) = 0.15 + 0.26667l_2 + 0.214285m_2 \quad (48)$$

and

$$C_2(1+m_2) = 0.029365 + 0.0762l_2 + 0.07626m_2 + 0.0309l_2^2 + 0.07619l_2m_2 + 0.0401m_2^2. \quad (49)$$

3.5. Method of solution

For convenience we now introduce dimensionless boundary layer thicknesses ξ_1 and ξ_2 such that

$$\xi_1 = \delta_1 \left(\frac{2\nu_1 x}{U_1} \right)^{-\frac{1}{2}}$$

$$\xi_2 = \delta_2 \left(\frac{2\nu_2 x}{U_1} \right)^{-\frac{1}{2}},$$

and put

$$a_1 = \frac{1 - a_0}{A_1 \xi_1}$$

$$a_2 = \frac{\lambda - a_0}{A_2 \xi_2}.$$

These definitions of a_1 and a_2 may be restated as

$$a_0 + a_1 A_1 \xi_1 = 1, \quad (50)$$

$$a_0 + a_2 A_2 \xi_2 = \lambda. \quad (51)$$

Equations (41), (44), (47) then become, respectively,

$$a_1 = - \left(\frac{\mu_2 \rho_2}{\mu_1 \rho_1} \right)^{\frac{1}{2}} \left(\frac{1 + m_2}{1 + m_1} \right) a_2, \quad (41a)$$

$$a_1^2 = B_1 a_0 a_1^2 \xi_1^2 + C_1 a_1^3 \xi_1^3, \quad (44a)$$

$$a_2^2 = B_2 a_0 a_2^2 \xi_2^2 + C_2 a_2^3 \xi_2^3. \quad (47a)$$

Equations (42) and (43), together with (36) and (38), become respectively

$$64l_1 = - \frac{4a_0(1-a_0)^2}{a_1^2} [(1+m_1)(1+\frac{16}{15}l_1+m_1)^{-2}], \quad (42a)$$

$$64l_2 = - \frac{4a_0(\lambda-a_0)^2}{a_2^2} [(1+m_2)(1+\frac{16}{15}l_2+m_2)^{-2}]. \quad (43a)$$

(45) These equations, together with (39) and (40), constitute nine equations to determine $a_0, a_1, a_2, \xi_1, \xi_2, l_1, m_1, l_2$, and m_2 .

(46) To obtain the first approximate solution the constants l_1, m_1, l_2 , and m_2 are put equal to zero. This gives the solution for the quartic polynomial velocity distribution—equations (33) and (34). The values of a_0, a_1, a_2, ξ_1 , and ξ_2 so obtained are then used to calculate values of l_1, m_1, l_2 , and m_2 to be used in the second approximation. In evaluating l_1 and l_2 from equations (42 a) and (43 a) squares and cubes of the l 's and m 's can be neglected.

(48) 3.6. Boundary layer properties

(49) The properties of interest are the interface velocity u_0 which is given directly by the solution of the equations (this quantity is highly important for the mass transfer solutions since it determines the time of contact), and the skin friction coefficient c_{τ_0} .

If the tangential stress at the interface is τ_0 , then

$$\tau_0 = \mu_1 \left(\frac{\partial u}{\partial y} \right)_{y=0+} = \frac{a_1(1+m_1)}{\sqrt{2}} \rho_1 U_1^2 \left(\frac{U_1 x}{v_1} \right)^{-\frac{1}{2}},$$

i.e.
$$\frac{\tau_0}{\rho_1 U_1^2} = c_{\tau_0} \left(\frac{U_1 x}{v_1} \right)^{-\frac{1}{2}},$$

where
$$c_{\tau_0} = \frac{a_1(1+m_1)}{\sqrt{2}}.$$

4. Results

Calculations have been carried out for the cases

$$\lambda = 0.501 \quad \text{and} \quad \mu_2 \rho_2 / (\mu_1 \rho_1) = 1,$$

(50) also for $\lambda = 10$ and $\mu_2 \rho_2 / (\mu_1 \rho_1) = 0.01, 0.10, 0.3162, 1.0, 3.162, 10, 100$.
(51) (In liquid-liquid interaction the values of $\mu_2 \rho_2 / (\mu_1 \rho_1)$ of interest lie in the region about 1.) The results of these calculations are presented in Tables 1 and 2.

(41 a) The first calculations made for this system are those of Keulegan (1) who presented the results of calculations for quartic and sextic polynomial velocity distributions for values of $\lambda = 0$ and $\mu_2 \rho_2 / (\mu_1 \rho_1) = 0.01, 0.10,$
(44 a) $0.3162, 1.0, 3.162, 10.0, 100$.

(47 a) Lessen (3) appears to have obtained an exact solution for

$$\lambda = 0, \quad \mu_2 \rho_2 / (\mu_1 \rho_1) = 1,$$

(42 a) while Lock (2) has calculated exact solutions and approximate solutions (quartic polynomial or sine velocity distributions) for

$$\lambda = 0, \quad \mu_2 \rho_2 / (\mu_1 \rho_1) = 5.965 \times 10^4, 100, 10, 1,$$

(43 a) and for $\lambda = 0.501, \mu_2 \rho_2 / (\mu_1 \rho_1) = 1$.

TABLE 1
Table of constants l_1 , l_2 and m_1 , m_2

λ	$\frac{\mu_2 \rho_2}{\mu_1 \rho_1}$	$-l_1$	$-m_1$	$-l_2$	$-m_2$
0.501	1.0	0.1124	0.0401	0.1408	0.0249
10.0	0.01	0.1534	0.0182	0.0568	0.0697
10.0	0.10	0.1661	0.0114	0.0794	0.0576
10.0	0.3162	0.1707	0.0089	0.0912	0.0513
10.0	1.00	0.1741	0.0072	0.1016	0.0458
10.0	3.162	0.1763	0.0060	0.1096	0.0416
10.0	10.0	0.1777	0.0052	0.1153	0.0385
10.0	100.0	0.1790	0.0045	0.1211	0.0354

TABLE 2
Boundary layer constants for the sextic polynomial solution

λ	$\frac{\mu_2 \rho_2}{\mu_1 \rho_1}$	a_0	$-a_1$	a_2	ξ_1	ξ_2
0.501	1.0	0.7660	0.1758	0.1731	3.1668	4.5555
10.0	0.01	2.2990	1.3865	14.634	2.2901	1.2100
10.0	0.10	3.8820	3.7903	12.573	1.8742	1.1348
10.0	0.3162	4.9809	5.8197	10.812	1.6910	1.0905
10.0	1.00	6.1699	8.3050	8.6411	1.5424	1.0481
10.0	3.162	7.2932	10.894	6.3541	1.4334	1.0123
10.0	10.0	8.2170	13.188	4.3150	1.3593	0.9855
10.0	100.0	9.3342	16.156	1.6672	1.2823	0.9561

TABLE 3
Interface velocities (u_0/U_1)

λ	$\frac{\mu_2 \rho_2}{\mu_1 \rho_1}$	Accurate solution	Quartic polynomial	Sextic polynomial
0	5.965×10^4	0.0211	0.0220	..
0	100	0.1727	0.1770	0.1739
0	10	0.3428	0.3480	0.3446
0	1	0.5873	0.5907	0.5888
0.501	1	0.7657	0.7668	0.7660

TABLE 4
Tangential stress coefficients (c_{τ_0})

λ	$\frac{\mu_2 \rho_2}{\mu_1 \rho_1}$	Accurate solution	Quartic polynomial	Sextic polynomial
0	5.965×10^4	0.3318	0.3309	..
0	100	0.3183	0.3186	0.3144
0	10	0.2815	0.2774	0.2772
0	1	0.1996	0.1943	0.1959
0.501	1	0.1219	0.1184	0.1212

In Tables 3 and 4 the various methods of solution are compared as far as existing calculations allow. It will be seen that the interface velocity given by the sextic polynomial is in every case better than that given by the quartic polynomial but this is not always the case for the skin friction coefficient c_{τ_0} .

5. Acknowledgements

The author wishes to acknowledge his debt to J. A. Storrow for suggesting the problem of liquid extraction across plane interfaces and for his constant stimulus and encouragement, and also to Miss R. Rogers, R. W. Maxwell, and G. A. Turner. The advice of the referees has been of great assistance in preparing this paper for publication.

REFERENCES

1. G. H. KEULEGAN, 'Laminar flow at the interface of two liquids', *J. Res. Nat. Bur. Std. (U.S.A.)*, **32** (1944) 303.
2. R. C. LOCK, 'The velocity distribution in the laminar boundary layer between parallel streams', *Quart. J. Mech. App. Math.* **4** (1951) 42.
3. M. LESSEN, 'On the stability of the free laminar boundary layer between parallel streams', N.A.C.A. Tech. Note 1929, August 1949 (unpublished).
4. C. S. YIH, 'Stability of parallel laminar flow with a free surface', *Proc. 2nd U.S. Nat. Cong. App. Mech.* 1954, p. 623.

THE FLOW OF FLUID ALONG CYLINDERS

By J. C. COOKE. (*University of Malaya, Cluny Road, Singapore*)

[Received 16 August 1956]

SUMMARY

The boundary layer equations for uniform flow parallel to the generators of any cylinder without corners are put into the form of a series of linear third-order differential equations. The first three of these are the same as those obtained by Seban and Bond (1) for a circular cylinder and solved by Kelly (2). The rest have additional terms depending on the radius of curvature of the cylinder and its derivatives. The problem is also attacked by a Pohlhausen method as far as four terms of the series. For large distances from the front, Rayleigh's method, as given by Hasimoto (3), gives the first two terms of an asymptotic expansion for the drag. Explicit calculations are made of the drag of an elliptic cylinder of eccentricity $\frac{1}{2}\sqrt{3}$. There is evidence that the drag is everywhere less than that of a circular cylinder of the same perimeter.

1. Introduction

WE consider the flow of an incompressible fluid past a cylinder immersed in it, the fluid moving with velocity U parallel to the generators at infinity. Seban and Bond (1) considered the case of a circular cylinder and showed that, provided that the radius of the cylinder is large compared with the width of the boundary layer, the flow is the same as that for a flat plate; this is usually called Blasius flow. [See, for instance, Goldstein (4).] Since the boundary layer thickens as we go downstream, Blasius flow applies only for a short length of cylinder. By means of a suitable series, Seban and Bond extended the distance of applicability along the circular cylinder; a similar attack on the problem has been made by Sowerby and Cooke (5). Glauert and Lighthill (6) obtained a Pohlhausen solution and also extended the distance still further downstream. They showed that the Pohlhausen solution is reasonably accurate.

This paper treats the problem of a general cylinder which has no sudden changes of curvature along its cross-section boundary. The problem of the boundary layer along a wedge remains unsolved. The recent attempt by Loitsianskii and Bolshakov (7) only applies to the flow along the inside of a wedge; the problem appears to be one of remarkable difficulty.

We give the solution in the form of a series, the first three terms of which are the same as for a circular cylinder of radius equal to the local radius of curvature. It was, of course, to be expected that the first term would be the same (corresponding to the Blasius flow), and perhaps also the second. The succeeding terms consist of those for a circular cylinder plus certain extra terms which depend on the radius of curvature of the cylinder

and its derivatives. By the numerical solution of a series of third-order differential equations general results in terms of certain parameters could be given. Though the equations are linear, apart from the first one, the solution is laborious and is not attempted here.

However, a Pohlhausen method similar to the one employed by Glauert and Lighthill can be applied to the problem to give results in the form of a series. It is not easy to gauge the accuracy of the results by this method, but the error may be expected to be much the same as that suggested by Glauert and Lighthill in their solution of the problem of the flow along a circular cylinder. Results are applied to an elliptic cylinder, and there appears to be some evidence that the drag may be expected to be everywhere less than that of a circular cylinder of the same perimeter.

The important results are that the first three terms of the Seban and Bond series, and of the Glauert and Lighthill series, for the local skin friction of a circular cylinder of radius a apply to a general cylinder if the radius a is replaced by the local radius of curvature ρ . Indeed, if the total drag per unit length of a closed general cylinder is required, it is possible to make use of four terms of the series for a circular cylinder. It must be emphasized, however, that the general cylinder must have no sharp edges or sudden changes in curvature.

2. The equations of motion

The equation for the steady flow of a fluid is, in the usual notation,

$$\mathbf{q} \times \boldsymbol{\omega} = \text{grad}\left\{\left(p/\rho_0\right) + \frac{1}{2}q^2\right\} + \nu \text{curl} \boldsymbol{\omega},$$

where $\boldsymbol{\omega} = \text{curl} \mathbf{q}$, \mathbf{q} is the velocity and ρ_0 the density.

We take an orthogonal coordinate system x, y, θ , where x is distance measured along the axis of the cylinder, the planes $x = \text{constant}$ being right cross-sections; θ is the angle which a plane through a generator normal to the cylinder makes with some fixed plane belonging to the family, and y is distance along a normal to the cylinder. The cylinder thus belongs to a family of parallel surfaces, $y = \text{constant}$ and we may take $y = 0$ to be the given cylinder. If in the section by the plane $x = \text{constant}$ the radius of curvature of the curve $y = 0$ is ρ , then in general we can take the square of the line element to be

$$ds^2 = dx^2 + dy^2 + (\rho + y)^2 d\theta^2.$$

We shall denote the three components of velocity corresponding to the three coordinates by u, v, w , and will assume that $w = 0$, so that the fluid flows along the cylinder without cross-flow. We shall also write $r = \rho + y$; ρ is clearly a function of θ only.

We may now derive the equations of motion, making the usual boundary

layer approximations, but assuming that r is the same order as δ , instead of order unity, as is usual. We shall also suppose that ρ'/ρ is of order unity, dashes denoting differentiation with respect to θ . We then obtain

$$u \frac{\partial u}{\partial x} + v \frac{\partial u}{\partial y} = -\frac{1}{\rho_0} \frac{\partial p}{\partial x} + \frac{v}{r} \left(\frac{\partial}{\partial y} \left(r \frac{\partial u}{\partial y} \right) + \frac{\partial}{\partial \theta} \left(\frac{1}{r} \frac{\partial u}{\partial \theta} \right) \right),$$

$$O(\delta) = -\frac{1}{\rho_0} \frac{\partial p}{\partial y},$$

$$O(\delta) = -\frac{1}{r \rho_0} \frac{\partial p}{\partial \theta},$$

together with the equation of continuity

$$\frac{\partial}{\partial x}(ru) + \frac{\partial}{\partial y}(rv) = 0.$$

Here δ is the boundary layer thickness.

The second equation shows in the usual way that the pressure may be taken as constant across the boundary layer. We therefore take the pressure in the boundary layer to be the same as in the main stream, which in our case is constant. The third equation shows that in order for the flow to remain 'straight', i.e. $w = 0$, a cross pressure gradient is required. However, since the total change in pressure right round the cylinder is required to be of order δ^2 , our assumption $w = 0$ is justified when the pressure is in fact uniform.

Thus the equations to be solved, with uniform pressure, are

$$u \frac{\partial u}{\partial x} + v \frac{\partial u}{\partial y} = \frac{v}{r} \frac{\partial}{\partial y} \left(r \frac{\partial u}{\partial y} \right) + \frac{v}{r} \frac{\partial}{\partial \theta} \left(\frac{1}{r} \frac{\partial u}{\partial \theta} \right), \quad (1)$$

$$\frac{\partial}{\partial x}(ru) + \frac{\partial}{\partial y}(rv) = 0, \quad (2)$$

where $r = \rho + y$ and ρ is a known function of θ only.

For flow past a circular cylinder the equations are of the same form except for the last term in equation (1). This term thus represents the effect of the departure of the cylinder from the circular form.

3. The solution of the equations

We can satisfy equation (2) by writing

$$u = \frac{1}{r} \frac{\partial \psi}{\partial y}, \quad v = -\frac{1}{r} \frac{\partial \psi}{\partial x}.$$

, instead
er unity,
n

We write

$$\xi = \frac{4}{\rho} \left(\frac{vx}{U} \right)^{\frac{1}{2}}, \quad \eta = \frac{1}{4\rho} \left(\frac{U}{vx} \right)^{\frac{1}{2}} (r^2 - \rho^2),$$

$$\psi = \rho(vxU)^{\frac{1}{2}} \{f_0(\eta, \theta) + \xi f_1(\eta, \theta) + \xi^2 f_2(\eta, \theta) + \dots\}.$$

The velocity in the mainstream is U .

We shall adopt as being better known the substitutions of Seban and Bond (1) rather than those of Sowerby and Cooke (5). They differ only by numerical constants and certain signs. We shall take η and $(U/vx)^{\frac{1}{2}}$ to be positive always; ρ and ξ will then be positive or negative according as $|r|$ is greater or less than $|\rho|$, that is, according as the surface of the cylinder is convex or concave as seen from the fluid. The substitutions lead to the equation

$$u = \frac{1}{2}U(f'_0 + \xi f'_1 + \xi^2 f'_2 + \dots),$$

dashes denoting partial differentiation with respect to η . The boundary conditions are $u = v = 0$ at $\eta = 0$, and $u = U$, $v = 0$ at $\eta = \infty$, for all ξ and θ . These become

$$f_s(0) = f'_s(0) = 0; \quad f'_0(\infty) = 2, \quad f'_{s+1}(\infty) = 0$$

for $s = 0, 1, 2, \dots$

We express u and v in powers of ξ , substitute in equations (1) and (2) and equate coefficients of powers of ξ , making use of the expansions of ρ/r and r/ρ in powers of $\xi\eta$ from the relation

$$\frac{\rho}{r} = (1 + \xi\eta)^{\frac{1}{2}};$$

this expansion will be valid as long as $\xi\eta$ is less than unity, that is, $y < 0.414\rho$; thus the boundary layer thickness should not exceed 0.414 times the radius of curvature. With this proviso we obtain the following equations:

$$f'''_0 + f_0 f''_0 = 0, \quad (3)$$

$$f'''_1 + f_0 f''_1 - f'_0 f'_1 + 2f''_0 f_1 = -f''_0 - \eta f'''_0, \quad (4)$$

$$f'''_2 + f_0 f''_2 - 2f'_0 f'_2 + 3f''_0 f_2 = -\eta f'''_1 - f''_1 + f_1'^2 - 2f_1 f''_1, \quad (5)$$

$$f'''_3 + f_0 f''_3 - 3f'_0 f'_3 + 4f''_0 f_3 = -\eta f'''_2 - f''_2 + 3f_1 f'_2 - 3f_2 f'_1 - 2f_1 f''_2 + A(\eta^2 f''_0 + 4f'_1), \quad (6)$$

.

where

$$A = \frac{1}{16} \left(\frac{\rho''}{\rho} - \frac{3\rho'^2}{\rho^2} \right).$$

Here primes attached to ρ represent differentiations with respect to θ , whilst primes attached to any f represent partial differentiations with respect to η .

Equation (5) should contain terms involving partial derivatives of f_1 with respect to θ . However, since equation (3) does not contain θ explicitly and since it must satisfy the boundary conditions for all θ , f_0 will be independent of θ , and so its partial derivatives with respect to that variable do not appear in equation (5). From equations (4) and (5) the functions f_1 and f_2 are in the same way successively seen to be dependent only on η . Equation (6) will contain partial derivatives with respect to θ of earlier functions which also vanish. Higher order equations will be considerably more complicated, as they will involve non-vanishing partial derivatives with respect to θ of some earlier functions.

It will be seen that the first three of these equations are the same as those found by Seban and Bond for a circular cylinder. The fourth and later equations are the circular cylinder equations with extra terms depending on ρ and its derivatives on the right-hand side. A universal solution of equation (6) is $g_3(\eta) + Ah_3(\eta)$ where

$$\hbar(D)g_3 = -\eta f_2''' - f_2'' + 3f_1' f_2' - 3f_2 f_1'' - 2f_1 f_2'',$$

$$\hbar(D)h_3 = \eta^2 f_1' + 4f_1',$$

$$\hbar(D) \equiv D^3 + f_0 D^2 - 3f_0' D + 4f_0'', \quad D \equiv \partial/\partial\eta.$$

Each solution is to satisfy the same boundary conditions as f_3 . Similar methods may be used for the higher equations.

Because of the difficulty of solving the equations of Seban and Bond by desk machines, no attempt will be made here to solve them. Sowerby and Cooke found some instability in the solution of equation (4) by step-by-step methods. They obtained 0.692 for the value of $f_1'(0)$; Seban and Bond gave 0.704, and Kelly (2), using an analogue type of computer, gave 0.696, which we shall adopt. We see that there is some doubt about the correctness of even the second significant figure. Greater variations occur in the case of f_2 .

From the results we may find the skin friction τ given by

$$\tau = \mu(\partial u/\partial y)_{y=0};$$

we obtain

$$\begin{aligned} \frac{\rho\tau}{\mu U} &= \frac{1}{\xi} f_0''(0) + f_1''(0) + \xi f_2''(0) + \xi^2 \{g_3''(0) + Ah_3''(0)\} + \dots \\ &= 1.328\xi^{-1} + 0.696 - 0.199\xi + \xi^2 \{g_3''(0) + Ah_3''(0)\} + \dots \end{aligned} \quad (7)$$

using Kelly's results (2), as far as they were calculated.

Because of the difficulties mentioned relating to the solution, we shall attack the problem by Pohlhausen's method.

4. A Pohlhausen solution

Following the method of Glauert and Lighthill (6) we take

$$u = \frac{U}{\alpha} \log \left(1 + \frac{y}{\rho} \right) \quad (8)$$

as our distribution, where α is now a function of x and θ .

This form satisfies equation (1) at the wall of the cylinder, where $y = 0$, but not the equation obtained from equation (1) by differentiating it with respect to y and putting $y = 0$. It can be shown, however, that the unsatisfied terms will only be significant if the series which we shall obtain later for ϵ (see p. 319) is carried to higher powers than is done in this paper.

The edge of the boundary layer is given by $y = \delta$, $u = U$, where δ is the boundary layer thickness. Hence we have

$$\delta = \rho(e^\alpha - 1).$$

It will be noticed that this velocity distribution does not satisfy the condition $\partial u / \partial y = 0$ when $y = \delta$. Glauert and Lighthill consider it important to fulfil as many boundary conditions near the wall as possible, and their results justify this conjecture.

We next obtain a momentum integral equation. We do this by multiplying equation (1) by r and integrating with respect to y between the limits 0 and δ , making use of equation (2). The process is well known and given in Goldstein's book (4).

The result is

$$\frac{\partial}{\partial x} \int_0^\delta u(U-u)r \, dy = -\nu \left\{ r \frac{\partial u}{\partial y} \right\}_0^\delta - \frac{\nu U J}{\alpha}, \quad (9)$$

where

$$\frac{\nu U J}{\alpha} = \nu \int_0^\delta \frac{\partial}{\partial \theta} \left(\frac{1}{r} \frac{\partial u}{\partial \theta} \right) dy.$$

The first term in equation (9) has been shown by Glauert and Lighthill (6) to be

$$U^2 \rho^2 \frac{(2\alpha^2 - 3\alpha + 2)e^{2\alpha} - (\alpha + 2)}{4\alpha^3} \frac{\partial \alpha}{\partial x}. \quad (10)$$

The term $-\nu \{ r \partial u / \partial y \}_0^\delta$ is taken to vanish at the upper limit (as should be the case) though in fact it does not do so because, as has already been pointed out, the particular boundary condition $\partial u / \partial y = 0$ when $y = \delta$ is not satisfied by our form (8). At the lower limit the term becomes $\nu U / \alpha$.

The evaluation of J is straightforward. We have

$$\begin{aligned} & \frac{\alpha}{U} \frac{\partial}{\partial \theta} \left(\frac{1}{r} \frac{\partial u}{\partial \theta} \right) \\ &= (-a' + 2a\beta) \frac{y}{r^2} + 2a^2 \rho \frac{y}{r^3} + (\beta^2 - \beta') \frac{1}{r} \log \left(1 + \frac{y}{\rho} \right) + \frac{a\beta\rho}{r^2} \log \left(1 + \frac{y}{\rho} \right), \end{aligned}$$

where $a = \rho'/\rho$, $\beta = \alpha'/\alpha$, dashes denoting derivatives with respect to x and keeping x constant in the case of α and β .

Integrating this expression between the limits 0 and δ , that is, between 0 and $\rho(e^\alpha - 1)$, we have

$$\int \frac{y}{r^2} dy = -1 + \alpha + e^{-\alpha}, \quad \int \frac{\rho y}{r^3} dy = \frac{1}{2}(1 - e^{-\alpha})^2,$$

$$\int \frac{1}{r} \log\left(1 + \frac{y}{\rho}\right) dy = \frac{1}{2}\alpha^2, \quad \int \frac{\rho}{r^2} \log\left(1 + \frac{y}{\rho}\right) dy = 1 - e^{-\alpha} - \alpha e^{-\alpha},$$

and so

$$J = a\beta(-1 + 2\alpha + e^{-\alpha} - \alpha e^{-\alpha}) + a^2(1 - e^{-\alpha})^2 + \frac{1}{2}\alpha^2(\beta^2 - \beta') - a'(-1 + \alpha + e^{-\alpha}). \quad (11)$$

From equations (9) and (10), on multiplying by $4\alpha/\rho^2 U^2$, we have

$$\frac{(2\alpha^2 - 3\alpha + 2)e^{2\alpha} - (\alpha + 2)}{\alpha^2} \frac{\partial \alpha}{\partial x} = \frac{4\nu}{\rho^2 U} (1 - J), \quad (12)$$

where J is given by equation (11).

The solution of equation (12) is in general complicated, but it can be obtained as a series.† We note from equation (11) that J is of order α^2 . Hence we first put $J = 0$ in equation (12), obtaining as the first approximation the result of Glauert and Lighthill, viz.

$$\epsilon (= 12\nu x/\rho^2 U) = \alpha^2 + \frac{4}{3}\alpha^3 + \dots, \quad (13)$$

the series being correct as far as written if J is of order α^2 .

On differentiating the series with respect to θ and rearranging, we obtain

$$\frac{\partial \alpha}{\partial \theta} = -a\alpha + \frac{2a\alpha^2}{3} + \dots,$$

and since $\beta = \alpha'/\alpha$,

$$\beta = -a + \frac{2a\alpha}{3} + \dots,$$

$$\beta' = -a' + \frac{2a}{3}(a' - a^2) + \dots$$

Substituting these values in the expression (11) for J we find that the coefficient of α^2 in J vanishes and that the coefficient of α^3 is $\frac{1}{3}a^2 - \frac{1}{6}a'$. Consequently, as far as the term in α^3 , the value of J is $-8A\alpha^3/3$ in our previous notation. Hence J is of order α^3 and so we may write another term of the Glauert and Lighthill solution into the series for ϵ . Hence

$$\epsilon = \alpha^2 + \frac{4}{3}\alpha^3 + \frac{9}{10}\alpha^4 + a_5\alpha^5 + \dots, \quad (14)$$

† I am indebted to a referee for simplifying the original procedure adopted in solving this equation.

where a_5 is a function of θ to be determined from equation (12). Differentiating the series (14) with respect to x , using equation (13) and substituting in equation (12), we obtain

$$2\alpha + 4\alpha^2 + \frac{18}{5}\alpha^3 + \frac{32}{15}\alpha^4 + \dots = \left(1 + \frac{8A}{3}\alpha^3 + \dots\right) \left(2\alpha + 4\alpha^2 + \frac{18}{5}\alpha^3 + 5a_5\alpha^4 + \dots\right).$$

The coefficients of α , α^2 , and α^3 agree, as was to be expected, and we may determine a_5 by equating coefficients of α^4 . We obtain

$$\epsilon = \alpha^2 + \frac{4}{3}\alpha^3 + \frac{9}{10}\alpha^4 + \left(\frac{32}{75} - \frac{16A}{15}\right)\alpha^5 + \dots$$

This equation can be inverted to give

$$\frac{1}{\alpha} = \epsilon^{-1} + \frac{2}{3} - \frac{13}{60}\epsilon^{\frac{1}{2}} + \left(\frac{134}{675} - \frac{8A}{15}\right)\epsilon + \dots, \quad (15)$$

or, using our previous notation,

$$\frac{\rho\tau}{\mu U} = \alpha^{-1} = 1.188\xi^{-1} + 0.667 - 0.188\xi + (0.149 - 0.400A)\xi^2 + \dots,$$

which should be compared with equation (7). Glauert and Lighthill gave a series similar to (15) as far as the term in $\epsilon^{\frac{1}{2}}$. Their next term would have been $134\epsilon/675$, that is, it would not have included the part $-8A\epsilon/15$. Once more we note that the first three terms are the same as for a circular cylinder, but that there are differences in the fourth and succeeding terms.

5. Drag for small x

Since the drag per unit area is τ , the drag D per unit length is given by

$$D = \int \tau \rho d\theta,$$

the limits of integration being 0 and 2π for the whole of the outside of a closed cylinder. In general the drag is thus given by

$$D = \mu U \sum_{r=0}^{\infty} k^{r-1} J_r, \quad (16)$$

where $k^2 = 16\nu x/Ua^2$, a is a representative length, and

$$J_r = \int \left(\frac{a}{\rho}\right)^{r-1} f_r''(0) d\theta.$$

We may note that since $f_3'' = g_3'' + Ah_3''$, and

$$A = -16\rho^3 \frac{d^2}{ds^2} \left(\frac{1}{\rho}\right),$$

where $ds (= \rho d\theta)$ is the element of arc of the section, the value of J_3 for a closed cylinder reduces to

$$\int_0^{2\pi} \left(\frac{a}{\rho}\right)^2 g_3''(0) d\theta,$$

since the integral of the term involving $Ah_3''(0)$ vanishes.

6. Drag for large x

It is not so easy to find a satisfactory method of dealing with this region but we may note that Glauert and Lighthill (6) and Stewartson (8) have obtained (consistent) asymptotic solutions of the problem of the circular cylinder, and that these agree in their first and second terms with the drag given by the well-known Rayleigh method in which an infinite cylinder is impulsively started from rest with a velocity U parallel to its length. Batchelor (9) made the first attempt to solve the Rayleigh problem for the general cylinder and Hasimoto (3, 10) considerably improved on this. We therefore take the first two terms of Hasimoto's solution in the expectation that they may be applied to give a correct result for the drag in our boundary layer problem as far as they go.

Hasimoto effectively transforms conformally the section of the general cylinder into a circle whose radius c is chosen in such a way that the two regions at infinity are unchanged. Batchelor calls the cylinder having this section the 'equivalent circular cylinder' and he gives the first term for the drag. In fact, from Hasimoto's work we now know that he could have taken the first two terms of the well-known formula for the drag of a circular cylinder (Jaeger (11)) and written down the result for the drag D per unit length in the form

$$\frac{D}{4\pi\mu U} = \frac{1}{\beta} - \frac{\gamma}{\beta^2},$$

where $\beta = \log(4vt/c^2) - 2\gamma$, γ is Euler's constant, and c is the radius of the equivalent circular cylinder.

Alternatively, keeping two terms in δ , we may write

$$\frac{D}{4\pi\mu U} = \frac{1}{\delta} + \frac{\gamma}{\delta^2}, \quad (17)$$

where

$$\delta = \log(4vt/c^2).$$

Finally, then, on the assumption that Rayleigh's problem may pass over into the boundary layer problem by writing $t = x/U$, equation (17) holds in the latter problem when we write

$$\delta = \log(4vx/Uc^2). \quad (18)$$

7. Drag

We will
elliptic c
thickness

The m
The resu

where k
the sam

It is t
values o

The fi
the fact
elliptic
at large
solution

For la
and b'
perimet
Hence

where e

When
(17) and
circular
for all e
true for

1. R. A.
2. H. R.
3. H. F.
4. S. G.
5. L. S.
6. M. E.
7. L. G.
8. K. S.
9. G. K.
10. H. F.
11. J. C.

7. Drag of an elliptic cylinder

We will illustrate these results by finding expressions for the drag of an elliptic cylinder of perimeter $2\pi a$ and eccentricity $\sqrt{3}/2$, that is, having a thickness ratio of $1/2$.

The numerical work involves the complete elliptic integrals E and K . The result is that

$$D = 2\pi\mu U\{k^{-1}f_0''(0) + f_1''(0) + 1.629kf_2''(0) + 3.576k^2g_3''(0) + \dots\},$$

where $k^2 = 16vx/Ua^2$, as compared with the drag of a circular cylinder of the same perimeter which is

$$D = 2\pi\mu U\{k^{-1}f_0''(0) + f_1''(0) + kf_2''(0) + k^2g_3''(0) + \dots\}.$$

It is to be understood, of course, that these formulae apply for small values of x .

The first two terms are the same in the two expansions, but in view of the fact that $f_2''(0)$ is negative the drag for small values of k is less for the elliptic cylinder than the circular cylinder. The matter is not so definite at larger values of k since $g_3''(0)$ appears to be positive from the Pohlhausen solution given in section 4.

For large values of x , if the cylinder has major and minor semi-axes a' and b' , the equivalent circular cylinder has a radius $\frac{1}{2}(a' + b')$, whilst the perimeter of the ellipse is $4Ea'$, which must be put equal to $2\pi a$.

Hence the radius c of the equivalent circular cylinder is given by

$$c = \frac{a\pi\{1 + (1 - e^2)^{\frac{1}{2}}\}}{4E},$$

where e is the eccentricity and E has modulus e .

When $e = \sqrt{3}/2$ this gives $c = 0.803a$ and consequently from equations (17) and (18) the drag of the elliptic cylinder is again less than that of the circular cylinder with the same perimeter. This can be shown to be true for all elliptic cylinders, and indeed it seems very likely that the result is true for all closed cylinders.

REFERENCES

1. R. A. SEBAN and R. BOND, *J. Aero. Sci.* **18** (1951) 671.
2. H. R. KELLY, *J. Aero. Sci.* **21** (1954) 634.
3. H. HASIMOTO, *Proc. Phys. Soc. Japan*, **9** (1954) 611.
4. S. GOLDSTEIN (ed.), *Modern Developments in Fluid Mechanics* (Oxford, 1938).
5. L. SOWERBY and J. C. COOKE, *Quart. J. Mech. App. Math.* **6** (1953) 50.
6. M. B. GLAUERT and M. J. LIGTHILL, *Proc. Roy. Soc. A*, **230** (1955) 188.
7. L. G. LOITSIANSKII and V. P. BOLSHAKOV, N.A.C.A. Tech. Mem. 1308 (1951).
8. K. STEWARTSON, *Quart. J. Mech. App. Math.* **13** (1955) 113.
9. G. K. BATCHELOR, *Quart. J. Mech. App. Math.* **7** (1954) 179.
10. H. HASIMOTO, *Proc. Phys. Soc. Japan*, **10** (1955) 397.
11. J. C. JAEGER, *Proc. Roy. Soc. Edinburgh, A*, **61** (1942) 223.

ELECTROMAGNETIC WAVES IN NEARLY PERIODIC STRUCTURES

By E. WILD

(Department of Mathematics, University of Manchester)

[Received 22 August 1956]

SUMMARY

The theory of simple harmonic electromagnetic waves in a nearly periodic structure of the type used in linear accelerators is discussed by an expansion in a series of appropriately defined transmission modes of the field in a unit cell of the structure. The possibility of the expansion being assumed, it is shown that the coefficients can be expressed as integrals of the field over the input or output apertures of the unit cell. The modification to the corresponding solution in the strictly periodic structure can be calculated by perturbation theory. In the first approximation the structure can be represented by a series of four terminal networks, and the voltages and currents in these can be defined in such a way that the relevant parameters vary smoothly in the neighbourhood of resonance.

1. Introduction

THE problem discussed in this paper is that of the calculation of simple harmonic electromagnetic fields in structures which are nearly periodic in the following sense. The type of periodic structure concerned consists of an infinite series of identical cavities, each with perfectly conducting walls and an input and an output aperture, connected together so that the output aperture of each is the input aperture of the next. The nearly periodic structure is an approximation to a finite length of such a structure in which adjacent cavities are nearly, but not quite, identical; the difference between cavities separated by a large number of others may be large. The apertures of all the cavities, however, must be exactly congruent.

Consideration of this problem is suggested by the use of such structures in linear accelerators for charged particles, which has been extensively developed in recent years (see, for example, Slater (1), Alvarez *et al.* (2)). The present investigation was inspired, in particular, by the accelerator developed by Alvarez *et al.* (2), but is not confined to this case. The results are also applicable to an electrical transmission system consisting of a series of $2n$ -terminal networks, half the terminals of which can be considered as input terminals and half as output terminals. Such a network is, in fact, with suitable geometrical design and disposition of the circuit elements, a particular case of a structure of the type considered, the operating frequency being low enough for the equations of alternating current theory to provide an adequate description of the electromagnetic behaviour.

The characteristic property of a periodic structure of the type considered,

as of all periodic oscillating structures (Brillouin (3)), is their property of sustaining oscillations in which the oscillatory field has the same form, but possibly different amplitude and phase, in all the cavities, the amplitude ratio and phase difference between all pairs of consecutive cavities being the same. Such oscillations will be called transmission modes. Each transmission mode, considered from the point of view of a single cavity of the structure, is an electromagnetic field which has the same boundary values for its tangential components at the output aperture as at the input aperture, except for a difference of amplitude and phase. Thus one can speak of a transmission mode of a cavity as well as of a periodic structure provided that input and output apertures are congruent. (It is this which necessitates the condition of exact congruence of all the apertures in the nearly periodic structures under discussion.)

In the designs which have been used in practice the operating conditions are such that only one mode, or more precisely one pair of reciprocal modes in the sense defined in sections 2 and 3, is propagated without attenuation. In a strictly periodic structure each of these two modes would have the same amplitude in all cavities and the same phase difference between any pair of successive cavities. In a nearly periodic structure in favourable circumstances only such dominant modes will be effectively excited, but it will still be necessary to calculate their amplitude and phase in the different cavities as in a structure with many cavities these may differ significantly from the unperturbed values.

This calculation may be considered as part of a first-order perturbation theory approximation to an exact calculation of the field in the structure. Estimates of this approximate solution for the Alvarez accelerator have been made by a precisely formulated analogy with the effect of a small variation of diameter of an empty cylindrical waveguide (2). The problem has also been discussed by an analogy with one-dimensional wave propagation in a continuous medium with slowly varying propagation constant (Wilkins (4)).

The method adopted in the present paper is an expansion of the field in a cavity, or rather of the tangential components of this field in the input and output apertures, in a series of transmission mode fields. The possibility of such an expansion being assumed, it is shown that the coefficients of the expansion at the output aperture for a given field there can be expressed as integrals over the output aperture, and that the coefficients at the input aperture can be easily determined from them. Hence the field at the input aperture can be determined; and since this input aperture is the output aperture of the preceding cavity, by repetition the fields at the first input aperture can be determined from those at the last output aperture.

The expansion in modes is valid only if the aperture components of all the modes together form a closed set for electromagnetic fields in the aperture. This is known to be so in certain cases, and it is assumed true for the cavities considered. The proof of the existence in general of the reciprocal modes which are used in determining the coefficients in the expansion also depends on this assumption of closure. But in the case of unattenuated modes in all cavities, and all modes in symmetrical cavities (except for certain critical cases), the existence of the reciprocal mode can be shown directly. When the reciprocal mode does exist the first-order perturbation theory approximation to the solution mentioned above is valid independently of the assumption of closure and of the validity of the expansion.

By an appropriate transformation the relations between input and output fields for a single cavity can be expressed in terms of conventionally, but precisely, defined mode voltages and currents, so that each cavity is equivalent to an infinite set of independent four-terminal networks. In this notation, in the first approximation solution, the structure is represented by a single series of four terminal networks, one for each cavity and one for each junction between cavities. For asymmetrical cavities the voltage-current representation is rather complicated, and the mode amplitude representation is better. For symmetrical cavities the voltage-current relations become simpler, and the junction networks reduce to pure transformers.

The voltage-current representation is most valuable for the critical case in a symmetrical cavity, in which the dominant mode in the cavity is a resonant mode with zero tangential electric, or magnetic, field in the apertures. The mode of operation of the Alvarez accelerator is of this type. In this case the dominant pair of reciprocal modes coalesces in the single resonant mode, and, since the tangential electric or magnetic field in the aperture is zero, the aperture fields of the modes no longer form a closed set. This necessitates a modification in the mode expansion to replace the missing term, which results in a discontinuity in the mode amplitude representation at resonance. The voltage-current representation, however, can be defined in such a way that all the relevant parameters vary smoothly in the neighbourhood of resonance, so that calculations are made much easier.

The present paper is confined to investigation and development of the general method and is of quite general application. It seems, however, to provide a practicable method for computation, particularly of resonant structures such as the Alvarez resonator. An example of application to a practical structure is given in section 7 to illustrate this, and further investigations are proceeding.

2. Transmission modes in a cavity

Consider a cavity which consists of a finite empty region D , bounded partly by a surface S' occupied by perfectly conducting material and partly by two apertures S_a and S_b , the input and output apertures. The positive normal to the bounding surface is taken directed outwards from D on S' and S_b , and inwards to D on S_a , and the apertures are to be congruent in the sense that there is a rigid displacement $P \rightarrow P'$ which carries S_a into S_b and preserves the sense of the positive normal. In virtue of this congruence the cavity can form one of an infinite series of congruent cavities, coupled together so that the output aperture of each is the input aperture of the next, the whole forming a periodic structure such that the displacement $P \rightarrow P'$ carries each cavity into the position occupied by the next.

The components $E e^{ickt}$, $H e^{ickt}$ of a simple harmonic electromagnetic field of wave number $k/2\pi$ in D satisfy Maxwell's equations

$$\nabla \wedge \mathbf{E} = -ikZ_0 \mathbf{H}, \quad \nabla \wedge \mathbf{H} = ikY_0 \mathbf{E}, \quad (2.1)$$

in D , and the boundary conditions

$$\mathbf{n} \wedge \mathbf{E} = 0 \quad \text{on } S', \quad (2.2)$$

where \mathbf{n} is the positive normal unit vector on the boundary of the cavity and $Z_0 = 1/Y_0$ is the characteristic impedance of free space. (Rationalized m.k.s. units are used.)

If (\mathbf{E}, \mathbf{H}) is an electromagnetic field which satisfies (2.1) in the interior of the extended structure formed by the repetition of the cavity, and also satisfies (2.2) on its boundaries, then the field $(\mathbf{E}', \mathbf{H}')$ obtained by subjecting (\mathbf{E}, \mathbf{H}) to the displacement $P \rightarrow P'$ is another. The whole set of such fields for a fixed value of k constitutes a linear vector space, and the equation

$$\mathbf{T}(\mathbf{E}, \mathbf{H}) = (\mathbf{E}', \mathbf{H}'), \quad (2.3)$$

where $\mathbf{E}'(P') = \mathbf{E}(P)$, $\mathbf{H}'(P') = \mathbf{H}(P)$, defines a linear transformation, \mathbf{T} , of this vector space into itself. An eigenvector of this transformation with eigenvalue m , that is, a field (\mathbf{E}, \mathbf{H}) such that

$$\mathbf{T}(\mathbf{E}, \mathbf{H}) = (m\mathbf{E}, m\mathbf{H}), \quad (2.4)$$

is a field which is the same in all the cavities of the extended structure except for multiplication by a factor $1/m$ in going from one cavity to the next. The corresponding field in the cavity D thus satisfies the further boundary conditions

$$\mathbf{E}(P) = m\mathbf{E}(P'), \quad \mathbf{H}(P) = m\mathbf{H}(P'), \quad (2.5)$$

where P is any point on S_a and P' is the corresponding point on S_b .

A field which satisfies (2.1), (2.2), and (2.5) will be called a transmission mode for the cavity, and m will be called the transmission factor of the mode.

In a vector space of finite dimension number every linear operator has at least one eigenvector. A $2n$ -terminal network with n input and n output terminals is a system with $2(n-1)$ degrees of freedom, since the independent variables can be taken as the $n-1$ independent input and output currents. Such a network therefore always has at least one transmission mode. If the dimension number is infinite, as is generally the case in the present application, there may be no eigenvectors, and it is therefore not obvious that all cavities of the type under consideration have transmission modes. However, such modes are known to exist in many cases of practical importance; an example additional to those of the linear accelerators mentioned in the introduction is that of a finite length of a cylindrical waveguide terminated by plane apertures perpendicular to the generators of the cylinder. The well known waveguide modes are transmission modes for such a cavity.

Besides the transmission modes there may, according to the general theory of linear operators, be associated with a given eigenvalue, m , of \mathbf{T} a set of fields $(\mathbf{E}_1, \mathbf{H}_1), (\mathbf{E}_2, \mathbf{H}_2), \dots, (\mathbf{E}_n, \mathbf{H}_n)$ such that

$$\left. \begin{aligned} \mathbf{T}(\mathbf{E}_1, \mathbf{H}_1) &= (m\mathbf{E}_1, m\mathbf{H}_1) \\ \mathbf{T}(\mathbf{E}_r, \mathbf{H}_r) &= (m\mathbf{E}_r + \mathbf{E}_{r-1}, m\mathbf{H}_r + \mathbf{H}_{r-1}) \quad \text{for } r = 2, 3, \dots, n \end{aligned} \right\} \quad (2.6)$$

Such a set of associated solutions, with $m = \pm 1$, $n = 2$, is of importance in the case of a symmetrical cavity operated at a resonance frequency. This case is discussed in section 6.

To every solution (\mathbf{E}, \mathbf{H}) of (2.1) there corresponds the conjugate solution $(\mathbf{E}^*, -\mathbf{H}^*)$, ($*$ denotes conjugate complex). If (\mathbf{E}, \mathbf{H}) is a transmission mode with transmission factor m , the conjugate solution is a transmission mode with factor m^* . If $|m| = 1$ then $m^* = 1/m$, and the conjugate mode may be considered as a mode with factor m in the sense from S_b to S_a , a mode 'reciprocal' to the original one. The existence of a reciprocal mode corresponding to any given mode in a symmetrical cavity (except for modes with $m = \pm 1$ in which the two modes in opposite sense may coincide) is obvious from symmetry. There is no such direct proof of the existence of reciprocal modes in asymmetrical cavities. It will be shown in section 3 to follow from the assumption of closure.

3. Orthogonality relations between transmission modes

If $(\mathbf{E}_1, \mathbf{H}_1)$ and $(\mathbf{E}_2, \mathbf{H}_2)$ are any two electromagnetic fields in D it follows from (2.1) that $\nabla \cdot (\mathbf{E}_1 \wedge \mathbf{H}_2 - \mathbf{E}_2 \wedge \mathbf{H}_1) = 0$. If the fields also satisfy the boundary condition (2.2) the use of the divergence theorem gives the equation

$$\int_{S_a} (\mathbf{E}_1 \cdot \tilde{\mathbf{H}}_2 - \mathbf{E}_2 \cdot \tilde{\mathbf{H}}_1) dS = \int_{S_b} (\mathbf{E}_1 \cdot \tilde{\mathbf{H}}_2 - \mathbf{E}_2 \cdot \tilde{\mathbf{H}}_1) dS, \quad (3.1)$$

where $\tilde{\mathbf{H}} = \mathbf{H} \wedge \mathbf{n}$. The integrals depend only on the tangential components of the fields in the aperture. If $(\mathbf{E}_1, \mathbf{H}_1)$ and $(\mathbf{E}_2, \mathbf{H}_2)$ are the fields of transmission modes with factors m_1 and m_2 respectively, (3.1) reduces to

$$(1 - m_1 m_2) \int_S (\mathbf{E}_1 \cdot \tilde{\mathbf{H}}_2 - \mathbf{E}_2 \cdot \tilde{\mathbf{H}}_1) dS = 0, \quad (3.2)$$

where S denotes either of the apertures S_a or S_b . The surface integral is therefore zero unless $m_2 = 1/m_1$.

Now consider the Hilbert space whose elements are pairs (\mathbf{A}, \mathbf{B}) of tangential vector functions defined on an aperture S congruent to S_a , with the scalar product definition

$$\{(\mathbf{A}_1, \mathbf{B}_1), (\mathbf{A}_2, \mathbf{B}_2)\} = \int_S (\mathbf{A}_1 \cdot \mathbf{A}_2^* + \mathbf{B}_1 \cdot \mathbf{B}_2^*) dS, \quad (3.3)$$

in which $\mathbf{A}_1 \cdot \mathbf{A}_2^*$ denotes the ordinary scalar product of cartesian vectors. (3.2) shows that the pairs of fields which consist of the tangential components in the aperture of $(\mathbf{E}_1, \tilde{\mathbf{H}}_1)$ and $(\tilde{\mathbf{H}}_2^*, -\mathbf{E}_2^*)$ are mutually orthogonal elements of this vector space if $m_2 \neq 1/m_1$. The result obviously holds for any value of m_1 if $\mathbf{E}_2 = \mathbf{E}_1$ and $\mathbf{H}_2 = \mathbf{H}_1$.

We now assume that the aperture tangential components of the fields of all transmission modes of a given wave number in the cavity form a complete set of elements of the Hilbert space; that is, that any pair of vector fields which is orthogonal to all of them is zero. This assumption is equivalent to the assumption of closure in the mean square sense. The investigation of the validity of this assumption for a given cavity would be difficult. It is known to be true for cylindrical waveguides (subject to restrictions on the smoothness of the boundary of the cross section) except at the critical frequencies for the modes, and it seems not unreasonable to assume that it will hold for any cavity of reasonably smooth shape except in certain critical conditions of the type discussed in section 6.

Applying the completeness assumption to the element $(\tilde{\mathbf{H}}^*, -\mathbf{E}^*)$ where \mathbf{E} and \mathbf{H} are the aperture tangential components of a transmission mode with factor m , we see that there must be a mode with aperture components not orthogonal to this element, that is, a mode with factor $1/m$; this is the reciprocal mode. The two modes must be distinct even if $m = \pm 1$, so that $m = 1/m$, since $(\tilde{\mathbf{H}}^*, -\mathbf{E}^*)$ is orthogonal to $(\mathbf{E}, \tilde{\mathbf{H}})$.

The transmission modes therefore occur in general in groups of four, each one being associated with conjugate, reciprocal, and conjugate reciprocal modes with factors m^* , $1/m$, and $1/m^*$ respectively. The four may reduce to two if m is real ($m = m^*$) or unimodular ($m = 1/m^*$) since

in those cases a mode may be identical with its conjugate or with its conjugate reciprocal. It has been stated (Slater (1)) that in the case of electromagnetic waves in lossless cavities m is always real or unimodular, but an example given in section 5 shows that this is not true.

The transmission modes can thus be grouped into pairs of reciprocal modes. The pairs will be numbered 1, 2, ..., and the members of the r th pair labelled $r+$ and $r-$ (the transmission factors being m_r and $1/m_r$ respectively). For each mode choose a definite amplitude as unit amplitude and denote the tangential electric and magnetic fields in the output aperture for this amplitude by \mathcal{E} and \mathcal{H} with the appropriate suffix. The unit amplitudes of conjugate modes are to be chosen so that $\mathcal{E}_{r\pm}^* = \mathcal{E}_{r\pm}$, $\mathcal{H}_{r\pm}^* = -\mathcal{H}_{r\pm}$ where $r'+$ and $r'-$ are the labels of the modes conjugate to $r+$ and $r-$ respectively. Then by (3.2) the two sets of elements $(\mathcal{E}_{1+}, \mathcal{H}_{1+}), (\mathcal{E}_{1-}, \mathcal{H}_{1-}), \dots$ and $(\mathcal{H}_{1-}^*, -\mathcal{E}_{1-}^*), (\mathcal{H}_{1+}^*, -\mathcal{E}_{1+}^*), \dots$ in that order are a pair of complete bi-orthogonal systems with respect to the scalar product (3.3). We define

$$\left. \begin{aligned} \int_S (\mathcal{E}_{r+} \cdot \tilde{\mathcal{H}}_{r-} - \tilde{\mathcal{H}}_{r+} \cdot \mathcal{E}_{r-}) dS &= N_r \\ \text{and have also from (3.2)} \\ \int_S (\mathcal{E}_{r\pm} \cdot \tilde{\mathcal{H}}_{s\pm} - \tilde{\mathcal{H}}_{r\pm} \cdot \mathcal{E}_{s\pm}) dS &= 0 \\ \int_S (\mathcal{E}_{r+} \cdot \tilde{\mathcal{H}}_{s-} - \tilde{\mathcal{H}}_{r+} \cdot \mathcal{E}_{s-}) dS &= 0 \quad \text{for } r \neq s \end{aligned} \right\} \quad (3.4)$$

Since the systems are complete N_r is not zero; the unit amplitudes can be chosen so as to make $N_r = 1$, but this normalization is not always convenient.

In view of the orthogonality relations any pair $(\mathbf{E}, \tilde{\mathbf{H}})$ of tangential vector fields in S can be expanded formally in a series of the form

$$\left. \begin{aligned} \mathbf{E} &= \sum_r (a_{r+} \mathcal{E}_{r+} + a_{r-} \mathcal{E}_{r-}) \\ \tilde{\mathbf{H}} &= \sum_r (a_{r+} \tilde{\mathcal{H}}_{r+} + a_{r-} \tilde{\mathcal{H}}_{r-}) \end{aligned} \right\} \quad (3.5)$$

where

$$a_{r\pm} = \pm \frac{1}{N_r} \int_S (\mathbf{E} \cdot \tilde{\mathcal{H}}_{r\mp} - \tilde{\mathbf{H}} \cdot \mathcal{E}_{r\mp}) dS. \quad (3.6)$$

It has been assumed for the sake of simplicity that there is only one linearly independent transmission mode for each transmission factor m_r (and only one pair of reciprocal modes for $m_r = \pm 1$ if these values occur at all). If there is more than one it can be shown without difficulty that

a linear basis can be chosen for the modes associated with each m_r in such a way that the orthogonality relations (3.4) still hold.

We now consider the expansion in a series of the type (3.5) of the aperture components of an arbitrary electromagnetic field (\mathbf{E} , \mathbf{H}) in the cavity, and in particular the relations between the amplitude coefficients $a_{r\pm,a}$ and $a_{r\pm,b}$ for the input and output aperture fields. The series (3.5) may in general have only formal significance, but under sufficiently restrictive conditions on \mathbf{E} and \mathbf{H} it will be convergent and will determine also an expansion of the field everywhere in the cavity as the sum of a series of transmission modes. The coefficients in this series will be the same as the coefficients in the series (3.5) for the aperture field in the output aperture S_b , and by the defining property of transmission modes it then follows that the amplitudes for the field at the input aperture will be given by

$$a_{r+,a} = m_r a_{r+,b}, \quad a_{r-,a} = a_{r-,b}/m_r. \quad (3.7)$$

Relations (3.7) can, however, be obtained directly for the amplitude coefficients as defined by (3.6) without reference to the expansion, and even without the completeness assumption, provided that the reciprocal mode exists and the normalizing integral N_r is not zero. In fact equation (3.1) with $(\mathbf{E}_1, \mathbf{H}_1)$ as the given field (\mathbf{E}, \mathbf{H}) and $(\mathbf{E}_2, \mathbf{H}_2)$ as the $r+$ mode of unit amplitude, with aperture components $(\mathcal{E}_{r+}, \mathcal{H}_{r+})$ on S_b and $(m_r \mathcal{E}_{r+}, m_r \mathcal{H}_{r+})$ on S_a gives immediately

$$m_r \int_{S_a} (\mathbf{E} \cdot \tilde{\mathcal{H}}_{r+} - \tilde{\mathbf{H}} \cdot \mathcal{E}_{r+}) dS = \int_{S_b} (\mathbf{E} \cdot \tilde{\mathcal{H}}_{r+} - \tilde{\mathbf{H}} \cdot \mathcal{E}_{r+}) dS,$$

or from (3.6)

$$m_r a_{r-,a} = a_{r-,b},$$

which is the second of (3.7). The first of (3.7) follows similarly by taking $(\mathbf{E}_2, \mathbf{H}_2)$ as the $r-$ mode of unit amplitude.

The amplitudes a_{r+} and a_{r-} of the aperture components of a pair of reciprocal modes can be expressed by means of a linear transformation in terms of conventionally defined voltages v_r and currents i_r , and the input-output relations then assume one of the standard forms for a four terminal network. The definitions are

$$v_r = a_{r+} e_{r+} + a_{r-} e_{r-}, \quad i_r = a_{r+} h_{r+} + a_{r-} h_{r-}, \quad (3.8)$$

where $e_{r\pm}$, $h_{r\pm}$ are arbitrary constants, which for later convenience are so chosen that $e_{r\pm} = e_{r\pm}^*$, $h_{r\pm} = -h_{r\pm}^*$, which may be interpreted as the voltages and currents in the r th mode for unit amplitude of the $r\pm$ modes in the aperture. (3.8) can be solved for the amplitudes to give

$$a_{r+} = \frac{v_r - Z_{r-} i_r}{h_{r+}(Z_{r+} - Z_{r-})}, \quad a_{r-} = -\frac{v_r - Z_{r+} i_r}{h_{r-}(Z_{r+} - Z_{r-})}, \quad (3.9)$$

where $Z_{r\pm} = e_{r\pm}/h_{r\pm}$ are the characteristic impedances of the $r+$ and $r-$ modes. The transfer relations for voltages and currents can be obtained from (3.9), (3.8), and (3.7) in the form

$$\left. \begin{aligned} v_{r,a} &= \frac{m_r Z_{r+} - Z_{r-}/m_r}{Z_{r+} - Z_{r-}} v_{r,b} - \frac{Z_{r+} Z_{r-} (m_r - 1/m_r)}{Z_{r+} - Z_{r-}} i_{r,b} \\ i_{r,a} &= \frac{m_r - 1/m_r}{Z_{r+} - Z_{r-}} v_{r,b} - \frac{m_r Z_{r-} - Z_{r+}/m_r}{Z_{r+} - Z_{r-}} i_{r,b} \end{aligned} \right\}. \quad (3.10)$$

The constants $e_{r\pm}$, $h_{r\pm}$ may be so chosen that $Z_{r+} = -Z_{r-} = Z_r$, and if this is done (3.10) simplifies to (3.19) below.

The power transmitted through the cavity by a given electromagnetic field can also be expressed simply in terms of the amplitudes or the voltages and currents. The mean rate of energy transfer P across the surface S in the positive sense by an electromagnetic field (\mathbf{E} , \mathbf{H}) is given by

$$P = \frac{1}{4} \int_S (\mathbf{E} \cdot \tilde{\mathbf{H}}^* + \mathbf{E}^* \cdot \tilde{\mathbf{H}}) dS. \quad (3.11)$$

Since the field conjugate to a transmission mode is also a transmission mode the aperture components of the field (\mathbf{E}^* , \mathbf{H}^*) can be expressed explicitly as a series of transmission modes in the form

$$\begin{aligned} \mathbf{E}^* &= \sum_r (a_{r+}^* \mathcal{E}_{r+} + a_{r-}^* \mathcal{E}_{r-}), \\ \tilde{\mathbf{H}}^* &= - \sum_r (a_{r+}^* \tilde{\mathcal{H}}_{r+} + a_{r-}^* \tilde{\mathcal{H}}_{r-}). \end{aligned}$$

Thus, by the orthogonality relations (3.4), (3.11) gives

$$P = \frac{1}{4} \sum_r N_r (a_{r+}^* a_{r-} - a_{r-}^* a_{r+}). \quad (3.12)$$

The various arbitrary constants have been chosen so that

$$\left. \begin{aligned} \mathcal{E}_{r'\pm} &= \mathcal{E}_{r\pm}^*, & \tilde{\mathcal{H}}_{r'\pm} &= -\tilde{\mathcal{H}}_{r\pm}^*, & e_{r'\pm} &= e_{r\pm}^*, & h_{r'\pm} &= -h_{r\pm}^* \end{aligned} \right\} \quad (3.13)$$

from which follow

$$Z_{r'\pm} = -Z_{r\pm}^*, \quad N_{r'} = -N_r^*$$

It can be verified that these conditions are compatible for modes which are self-conjugate or conjugate to their reciprocals. With them and (3.9), (3.12) can be transformed to

$$P = \frac{1}{4} \sum_r \{N_r (v_r i_{r'}^* + v_{r'}^* i_r) / (e_{r+} h_{r-} - e_{r-} h_{r+})\},$$

so that if the constants are chosen so that

$$N_r = e_{r+} h_{r-} - e_{r-} h_{r+}, \quad (3.14)$$

a choice which is compatible with (3.13), the expression for the power

transfer reduces to

$$P = \frac{1}{4} \sum_r (v_r i_r^* + v_r^* i_r).$$

For modes which are self-conjugate or conjugate to their reciprocals $v_r = v_r^*$, $i_r = i_r^*$ and the corresponding term is real and occurs only once in the summation. Every other term appears together with a complementary term $v_r^* i_r^* + v_r i_r$ which is conjugate to it. Thus the formula for power transfer can finally be written as

$$P = \frac{1}{4} \sum_r (v_r i_r^* + v_r^* i_r) = \frac{1}{2} \operatorname{re} \left(\sum_r v_r i_r^* \right), \quad (3.15)$$

a generalization of the formula $P = \frac{1}{2} \operatorname{re} \{vi^*\}$ for power input at a pair of terminals.

The introduction of the voltage-current notation does not simplify the expansion formulae in the general case. For a symmetrical cavity, however, i.e. a cavity which has a plane of symmetry such that S_a and S_b are images of each other in it in such a way that any point P' of S_b is the image of the corresponding point P of S_a , such a simplification is possible. The simplification arises from the fact that to any solution $E_x, E_y, E_z, H_x, H_y, H_z$ of Maxwell's equations there corresponds a 'reflected solution', $E'_x, E'_y, -E'_z, -H'_x, -H'_y, H'_z$, where $E'_x(x, y, z) = E_x(x, y, -z)$, etc. In a symmetrical cavity in which the plane of symmetry is taken as the plane $z = 0$, the reflected solution satisfies the boundary conditions (2.2) if the original solution does, and also satisfies (2.5) with constant $1/m$ if the original solution does so with constant m . Hence the reflected solution of a transmission mode can be taken as the reciprocal mode. (This is obvious if there is only one mode for a given value of m . It can be arranged by a suitable choice of the linear basis even if there is more than one.)

The fields and arbitrary constants of the reciprocal modes can therefore be chosen so that

$$\begin{aligned} \mathcal{E}_{r+} = \mathcal{E}_{r-} = \mathcal{E}_r, \quad \mathcal{H}_{r+} = -\mathcal{H}_{r-} = \mathcal{H}_r, \quad e_{r+} = e_{r-} = e_r, \\ h_{r+} = -h_{r-} = h_r, \quad Z_{r+} = -Z_{r-} = Z_r. \end{aligned}$$

The orthogonality relations (3.4) reduce to

$$\int_S \mathcal{E}_r \mathcal{H}_s^* dS = \begin{cases} 0 & \text{if } r \neq s, \\ -\frac{1}{2} N_r = e_r h_r & \text{if } s = r, \end{cases} \quad (3.16)$$

where e_r and h_r have been chosen so that N_r has the value (3.14).

The expansion relations (3.5) and (3.6) reduce to

$$\mathbf{E} = \sum_r \frac{v_r}{e_r} \mathcal{E}_r, \quad \mathbf{H} = \sum_r \frac{i_r}{h_r} \mathcal{H}_r, \quad (3.17)$$

$$v_r = \frac{1}{h_r} \int_S \mathbf{E} \cdot \mathcal{H}_r^* dS, \quad i_r = \frac{1}{e_r} \int_S \mathcal{E}_r \cdot \mathbf{H}^* dS, \quad (3.18)$$

and the transfer relations (3.10) to

$$\left. \begin{aligned} v_{r,a} &= \frac{1}{2}(m_r + 1/m_r)v_{r,b} + \frac{1}{2}Z_r(m_r - 1/m_r)i_{r,b} \\ i_{r,a} &= \frac{1}{2Z_r}(m_r - 1/m_r)v_{r,b} + \frac{1}{2}(m_r + 1/m_r)i_{r,b} \end{aligned} \right\} \quad (3.19)$$

4. Application to nearly periodic structures

The resolution of the aperture fields into modes can now be applied to the calculation of fields in a structure consisting of a succession of cavities, not necessarily congruent but having congruent apertures. Every aperture except the two terminating ones is the output aperture of one cavity and the input aperture of the next. If the input amplitudes for any cavity are known the aperture components of the fields can be calculated from (3.5), hence the output amplitudes for the preceding cavity from (3.6) and the input amplitudes from (3.7). Thus the amplitudes at the first input can be expressed as linear combinations of those at the last output; the linear equations relating the two sets of amplitudes are then to be solved subject to the restrictions imposed by the terminal boundary conditions. If all the cavities are symmetrical the relations between the mode expansions at the output of one cavity and at the input of the next can be expressed simply in terms of the mode voltages and currents; in fact if functions belonging to the first cavity are distinguished by accents, we have from (3.17) and (3.18)

$$\left. \begin{aligned} v'_{r,b} &= \sum_s c_{rs} v_{s,a}, & i'_{r,b} &= \sum_s d_{rs} i_{s,a}, \\ \text{where } c_{rs} &= \frac{1}{h'_r e_s} \int_S \mathcal{E}_s \cdot \mathcal{H}'_r dS, & d_{rs} &= \frac{1}{e'_r h_s} \int_S \mathcal{E}'_r \cdot \mathcal{H}_s dS \end{aligned} \right\} \quad (4.1)$$

The theory of this section is valid for a structure each of whose cavities is of arbitrary shape provided that all the apertures are congruent as stipulated, but its practical value would be small in the absence of further simplifying conditions. One such condition is that adjacent cavities should be nearly identical, differing from each other by a 'small' amount measurable by a small positive number ϵ , so that the whole forms a nearly periodic structure and the modes in adjacent cavities are closely related. In this case the coupling constants between different modes in two adjacent cavities (the constants c_{rs} , d_{rs} of (4.1) for $r \neq s$ in the symmetrical case) will be small quantities of order ϵ , and the methods of perturbation theory will be available for solution of the equations. If, as is usually the case in practice, the system is operated in such conditions that in zero order approximation only a single mode and its reciprocal are excited, the modification of the amplitudes of the dominant modes due to excitation of the other modes

will be of order ϵ^2 . For calculations of the dominant mode only, accurate to first order in ϵ , the presence of the other modes can be ignored, and the system is reduced effectively to a system of four terminal networks in cascade. The validity of this approximation is independent of the validity of the expansion in transmission modes. The modifications introduced by the departures from periodicity are (1) the coefficients in the voltage-current transfer equations of a single cavity ((3.10), (3.19)) have different values for different cavities, (2) the output amplitudes or voltages and currents of each cavity are not the same as the input amplitudes or voltages and currents of the next, but must be obtained from them by further transfer equations. In the symmetrical case these reduce by (4.1) to a pure transformer action,

$$v'_{r,b} = c_{rr} v_{r,a}, \quad i'_{r,b} = d_{rr} i_{r,a}, \quad (4.2)$$

where, as may be shown without difficulty, $c_{rr} d_{rr} - 1$ is of order ϵ^2 .

5. Properties of the transmission modes in symmetrical cavities

The determination of the transmission modes of a given cavity is a boundary value problem of a rather unusual type. Approximate solutions can be obtained by standard mathematical techniques for certain simple types of cavity, but there is no general method. However, some insight into the properties of the modes can be obtained by general methods, particularly for symmetrical cavities, and to do this is the aim of this section and the next.

If suitable boundary conditions are imposed on the field at the apertures S_a and S_b (in addition to the condition $\mathbf{n} \wedge \mathbf{E} = 0$ on S') there will be a set of resonance values k_1, k_2, \dots of k for which there are fields which satisfy these boundary conditions. In particular we consider 'open-circuit' resonances, with the boundary condition $\mathbf{n} \wedge \mathbf{H} = 0$, and 'closed circuit' resonances, with the condition $\mathbf{n} \wedge \mathbf{E} = 0$, on S_a and S_b .

Suppose that k is not one of the open-circuit resonance values. Then the fields in the cavity, and in particular the aperture electric fields, are completely determined by the aperture magnetic fields. The dependence of the total field on the aperture magnetic field is linear, and four linear operators \mathbf{S}_{aa} , \mathbf{S}_{ab} , \mathbf{S}_{ba} , and \mathbf{S}_{bb} can be defined such that the aperture tangential electric fields E_a and E_b are related to the magnetic fields H_a and H_b by the equations

$$iE_a = \mathbf{S}_{aa} \tilde{H}_a + \mathbf{S}_{ab} \tilde{H}_b, \quad iE_b = \mathbf{S}_{ba} \tilde{H}_a + \mathbf{S}_{bb} \tilde{H}_b. \quad (5.1)$$

Since $(\mathbf{E}^*, -\mathbf{H}^*)$ is an electromagnetic field in the cavity whenever (\mathbf{E}, \mathbf{H}) is one, the operators \mathbf{S} are real (that is, $\mathbf{S}\mathbf{H}$ is real if \mathbf{H} is real).

If in (3.1) $(\mathbf{E}_1, \mathbf{H}_1)$ is taken as the field with aperture magnetic field components $H_a = \mathbf{H}_1$, $H_b = 0$, and $(\mathbf{E}_2, \mathbf{H}_2)$ as the field with aperture

components $\mathbf{H}_a = \mathbf{H}_2^*$, $\mathbf{H}_b = 0$, the resulting equation can be written as

$$\int_S (\mathbf{S}_{aa} \tilde{\mathbf{H}}_1) \cdot \tilde{\mathbf{H}}_2^* dS = \int_S \tilde{\mathbf{H}}_1 \cdot (\mathbf{S}_{aa} \tilde{\mathbf{H}}_2)^* dS,$$

which shows that the operator \mathbf{S}_{aa} is adjoint to itself with respect to the scalar product $\int_S \tilde{\mathbf{H}}_1 \cdot \tilde{\mathbf{H}}_2^* dS$. Similarly \mathbf{S}_{bb} is adjoint to itself, and \mathbf{S}_{ab} is adjoint to $-\mathbf{S}_{ba}$.

For symmetric cavities the existence of the reflected solution corresponding to any given solution shows that $\mathbf{S}_{aa} = -\mathbf{S}_{bb}$ and $\mathbf{S}_{ab} = -\mathbf{S}_{ba}$; thus in this case \mathbf{S}_{ba} is adjoint to itself also.

By writing the boundary conditions (2.5) in terms of \mathcal{H} and the operators \mathbf{S} it follows that a necessary and sufficient condition that \mathcal{H} should be the aperture magnetic field of a transmission mode with transmission factor m is that \mathcal{H} should satisfy the equation

$$m\mathbf{S}_{aa}\tilde{\mathcal{H}} + \mathbf{S}_{ab}\tilde{\mathcal{H}} = m(m\mathbf{S}_{ba}\tilde{\mathcal{H}} + \mathbf{S}_{bb}\tilde{\mathcal{H}}). \quad (5.2)$$

For symmetrical cavities (if the case $m = 0$ is excluded) this reduces to

$$\mathbf{S}_{aa}\tilde{\mathcal{H}} = \frac{1}{2}(m+1/m)\mathbf{S}_{ba}\tilde{\mathcal{H}}. \quad (5.3)$$

Equation (5.3) has the same form as the eigenvalue equation for the resonant frequencies of a mechanical system. The orthogonality relations (3.16) can be derived from it directly. If either operator \mathbf{S}_{aa} or \mathbf{S}_{ba} is positive definite it also follows that the eigenvalue $\frac{1}{2}(m+1/m)$ must be real, and that therefore m must be real or unimodular, as can be seen by multiplying (5.3) by $\tilde{\mathcal{H}}^*$ and integrating over S ,

$$\int_S (\mathbf{S}_{aa}\tilde{\mathcal{H}}) \cdot \tilde{\mathcal{H}}^* dS = \frac{1}{2}(m+1/m) \int_S (\mathbf{S}_{ba}\tilde{\mathcal{H}}) \cdot \tilde{\mathcal{H}}^* dS.$$

The two integrals are real, and therefore $\frac{1}{2}(m+1/m)$ is real unless both the integrals are zero.

In the case of a four-terminal network, in which the operators \mathbf{S} reduce to pure numbers which are necessarily real, it is obvious that $(m+1/m)$ is always real; $(m+1/m)$ is also real for modes in a cylindrical waveguide.

The operators, however, are not in general positive definite, and complex values of $\frac{1}{2}(m+1/m)$ are therefore possible. An example, for a system with a finite number of degrees of freedom, is given by the symmetrical lossless six terminal transmission network of Fig. 1, in which the quantities X are the reactances of the circuit elements. For this system the operators \mathbf{S} can be calculated from the network equations. The equation for the eigenvalue $\lambda = \frac{1}{2}(m+1/m)$ is found to be

$$\lambda^2(X_1^2 + \alpha^2 X_3^2) - 2\lambda\alpha(1-\alpha)X_3^2 + \{(1-\alpha)^2 X_3^2 - X_1^2\} = 0,$$

where $\alpha = X_2/(2X_2 + X_1)$. This equation obviously has complex roots for sufficiently small values of α (small values of X_2) if $X_3^2 > X_1^2 > 0$. In fact the roots are complex if

$$(1 - 2\alpha)X_3^2 > X_1^2 > 0.$$

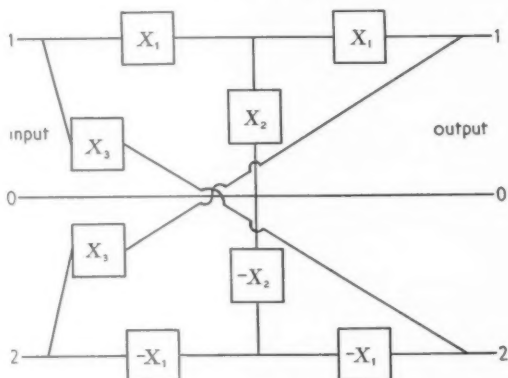


FIG. 1

The relation between the aperture electric and magnetic fields for a transmission mode in a symmetrical cavity can be put into a form which is more convenient for later developments. We have for such a mode

$$i\mathbf{E}_a = im\mathcal{E} = m\mathbf{S}_{aa}\mathcal{H} + \mathbf{S}_{ab}\tilde{\mathcal{H}},$$

hence

$$i\mathcal{E} = \mathbf{S}_{aa}\tilde{\mathcal{H}} + \frac{1}{m}\mathbf{S}_{ab}\mathcal{H},$$

but also

$$i\mathbf{E}_b = i\mathcal{E} = m\mathbf{S}_{ba}\mathcal{H} + \mathbf{S}_{bb}\tilde{\mathcal{H}},$$

and therefore from the relations $\mathbf{S}_{aa} = -\mathbf{S}_{bb}$, $\mathbf{S}_{ba} = -\mathbf{S}_{ab}$,

$$i\mathcal{E} = \frac{1}{2}(m-1/m)\mathbf{S}_{ba}\tilde{\mathcal{H}}. \quad (5.4)$$

This gives an alternative formula for the normalizing integral

$$\begin{aligned} e_r h_r &= \int_S \mathcal{E}_r \cdot \tilde{\mathcal{H}}_r dS \\ &= \frac{1}{2i} \left(m_r - \frac{1}{m_r} \right) \int_S (\mathbf{S}_{ba} \tilde{\mathcal{H}}_r) \cdot \tilde{\mathcal{H}}_r dS. \end{aligned} \quad (5.5)$$

6. Operation at and near resonance

In a symmetrical cavity, since the reflected solution corresponding to an open-circuit or closed-circuit resonant mode is another such mode,

these modes are, or in case of degeneracy a linear basis for them can be taken to be, all even or odd (that is, to have the non-vanishing aperture tangential components at the input and output apertures equal or equal and opposite). This being the case, such a resonant mode is also a transmission mode with transmission factor $+1$ for an even mode, -1 for an odd mode. Conversely, if there is a transmission mode with factor $m = \pm 1$ which is neither an open-circuit nor a closed-circuit resonant mode the sum and the difference of this mode and its corresponding reflected solution are open-circuit and closed-circuit resonant modes, even for $m = +1$, odd for $m = -1$. Thus transmission factors ± 1 occur at and only at resonant frequencies, and we wish to study the behaviour of these modes and of the modes which approach them in the neighbourhood of resonance.

Consider then a simple (non-degenerate) closed circuit even resonant mode. Two cases arise.

1. There is also an open-circuit even-resonant mode of the same frequency, and the aperture fields of the two modes are not orthogonal, $\int_S \mathbf{E} \cdot \tilde{\mathbf{H}} dS \neq 0$. Then any linear combination of the two modes is a transmission mode with factor $+1$; such a linear combination and the corresponding reflected solution form a pair of reciprocal modes to which the theory of section 3 can be applied directly.

2. There is no such open circuit mode. This case requires further consideration.

In case 2, let \mathcal{H}_0 be the tangential magnetic field in the input and output apertures for (arbitrarily chosen) unit amplitude of the resonant mode. Since there is no even open-circuit resonant mode there is a (unique) electromagnetic field in the cavity which has boundary values $\frac{1}{2}\mathcal{H}_0$ and $-\frac{1}{2}\mathcal{H}_0$ for the tangential magnetic field at the input and output apertures respectively, and by symmetry the tangential electric field for this mode will be the same, say \mathcal{E}_0 , at the input and output apertures. This field and the resonant mode obviously form a pair of associated solutions of the type described in equations (2.6). Since the resonant mode is a transmission mode with factor 1, \mathcal{H}_0 is orthogonal on S to the aperture tangential component \mathcal{E}_r ($r = 1, 2, \dots$) of the electrical field of every other transmission mode; and by applying (3.1) to the field of such a mode and the field with boundary values $(\mathcal{E}_0, \frac{1}{2}\mathcal{H}_0)$ and $(\mathcal{E}_0, -\frac{1}{2}\mathcal{H}_0)$ one can see that \mathcal{E}_0 is orthogonal to every aperture field $\tilde{\mathcal{H}}_r$. Thus the sets $(\mathcal{E}_0, \mathcal{E}_1, \dots)$ and $(\mathcal{H}_0, \mathcal{H}_1, \dots)$ are bi-orthogonal, and we assume them to be complete. (The set $(\mathcal{E}_1, \mathcal{E}_2, \dots)$ is obviously not complete since \mathcal{H}_0 is orthogonal to every member of it.)

We now put

$$\int_S \mathcal{E}_0 \cdot \mathcal{H}_0 dS = 1/Y, \quad (6.1)$$

and define the voltage v_0 and current i_0 of the 0th mode in a given electromagnetic field (\mathbf{E} , \mathbf{H}) at the aperture S by

$$v_0 = \int_S \mathbf{E} \cdot \tilde{\mathcal{H}}_0 dS, \quad i_0 = Y \int_S \mathcal{E}_0 \cdot \tilde{\mathbf{H}} dS, \quad (6.2)$$

so that the expansions (3.17) hold with $h_0 = 1$ and $e_0 = 1/Y$.

The transfer equations from output to input now become

$$v_{0,a} = v_{0,b}, \quad i_{0,a} = Y v_{0,b} + i_{0,b}. \quad (6.3)$$

Equations (6.3) are the limiting forms of equations (3.19), with $r = 0$, as m_0 tends to 1 and Z_0 tends to 0 in such a way that $\frac{1}{2}(m_0 - 1/m_0)/Z_0$ tends to Y . It will now be shown, with reasonable assumptions, that this limiting process is significant in that, for small variations of the shape and size of the cavity which involve a departure from resonance, there is a pair of reciprocal transmission modes whose voltage and current relations tend smoothly to those of (6.2) and (6.3) as resonance is approached. It is assumed for this purpose that there are no open-circuit resonant modes of either parity at the operating frequency.

The aperture field \mathcal{H} and the eigenvalue $\lambda = \frac{1}{2}(m+1/m)$ for a transmission mode are determined by equation (5.3)

$$\mathbf{S}'_{aa} \tilde{\mathcal{H}} = \lambda \mathbf{S}'_{ba} \tilde{\mathcal{H}},$$

(the ' indicating functions corresponding to the distorted cavity) and the operators \mathbf{S}'_{aa} and \mathbf{S}'_{ba} may be assumed to vary smoothly with the dimensions of the cavity since the resonant mode involved is a closed circuit, not an open circuit. For the undistorted cavity the equation has a non-degenerate solution $\mathcal{H} = \mathcal{H}_0$ with $\lambda = 1$, and it may be assumed that for small changes in dimensions there will be a solution \mathcal{H}_0 , with eigenvalue λ' , which by suitable choice of the arbitrary amplitude factor can be taken to tend smoothly to \mathcal{H}_0 , λ' tending smoothly to 1, as the changes in dimensions tend to zero. This solution gives rise to two reciprocal modes with the two solutions m'_0 and $1/m'_0$ of the equation $\frac{1}{2}(m+1/m) = \lambda'$ as transmission factors. We take for these modes $h'_0 = 1$ so that $e'_0 = Z'_0$ and, from (5.5),

$$Z'_0 = e'_0 = \frac{1}{2i} (m'_0 - 1/m'_0) \int_S (\mathbf{S}'_{ba} \tilde{\mathcal{H}}'_0) \cdot \tilde{\mathcal{H}}'_0 dS,$$

while from (5.4) the aperture electric field \mathcal{E}'_0 is given by

$$\mathcal{E}'_0 = \frac{1}{2i} (m'_0 - 1/m'_0) \mathbf{S}'_{ba} \tilde{\mathcal{H}}'_0.$$

In the undistorted cavity there can exist an electromagnetic field with boundary values $(\mathcal{E}_0, \mathcal{H}_0)$ at the input aperture and $(\mathcal{E}_0, 0)$ at the output aperture, this being the sum of the resonant mode, with boundary values

$(0, \frac{1}{2}\mathcal{H}_0)$, $(0, \frac{1}{2}\mathcal{H}_0)$, and the field with boundary values $(\mathcal{E}_0, \frac{1}{2}\mathcal{H}_0)$, $(\mathcal{E}_0, -\frac{1}{2}\mathcal{H}_0)$. Thus in the undistorted cavity $\mathbf{S}_{aa}\mathcal{H}_0 = \mathbf{S}_{ba}\mathcal{H}_0 = i\mathcal{E}_0$, and $\mathbf{S}'_{ba}\mathcal{H}'_0$ tends smoothly to $i\mathcal{E}_0$ as the distortion tends to zero.

Thus the functions

$$\frac{1}{2}(m'_0 + 1/m'_0),$$

$$\frac{1}{2}(m'_0 - 1/m'_0)Z'_0 = \frac{1}{i}(\lambda'^2 - 1) \int_S (\mathbf{S}'_{ba}\mathcal{H}'_0) \cdot \mathcal{H}'_0 dS, \quad \frac{1}{2}(m'_0 - 1/m'_0)/Z'_0,$$

\mathcal{H}'_0/h'_0 and \mathcal{E}'_0/e'_0 tend smoothly to the limits 1, 0, Y , \mathcal{H}_0 , and $Y\mathcal{E}_0$ respectively; the coefficients of the transfer equations and the voltage and current of a fixed aperture field tend smoothly to their values for the undistorted cavity.

In general m'_0 itself and $\frac{1}{2}(m'_0 - 1/m'_0)$ will not vary smoothly since if $\frac{1}{2}(m'_0 + 1/m'_0) = 1 + \epsilon$ with ϵ small, $m'_0 \sim 1 \pm \sqrt{2\epsilon}$. If ϵ is positive m'_0 is real, if ϵ is negative m'_0 is complex and unimodular. The above results apply to variations of the operating frequency as well as to distortions of the cavity, so that in general a resonance of this type separates a pass band from an attenuation band of the mode concerned.

The methods of this section can be applied to an odd closed-circuit resonance, and the results hold with small modifications in detail. For the discussion of open-circuit resonances a development parallel to that of section 5 with the roles of the electric and magnetic fields interchanged is necessary. Smooth variation of the voltage and current relations is then obtained by taking $e'_0 = 1$ so that h'_0 tends to 0.

The phenomenon of an open-circuit and a closed-circuit resonant mode of the same parity at the same frequency (case 1 above) occurs in a structure of n identical symmetrical cavities (considered as a single cavity) when each cavity has a transmission mode of factor m such that $m^n = \pm 1$. Such a value for the individual cavity is of no special significance and this suggests that case 1 in general is an accidental degeneracy of no particular importance. The transmission modes in the neighbourhood of such an even resonance can be shown to be determined by an eigenvalue equation

$$\mathbf{S}_1 \mathcal{E} = i \frac{m-1}{m+1} \mathcal{H}, \quad \mathbf{S}_2 \mathcal{H} = i \frac{m-1}{m+1} \mathcal{E},$$

where $\mathbf{S}_2 = \mathbf{S}_{aa} - \mathbf{S}_{ba}$ and \mathbf{S}_1 is similarly defined in terms of operators which give the aperture magnetic fields as functions of the electric fields. The eigenvalue m varies smoothly in the neighbourhood of $m = 1$ and for any one-parameter family of variations of \mathbf{S}_1 and \mathbf{S}_2 there is a pair of transmission modes which tend to definite linear combinations of the open-circuit and closed-circuit resonant modes. This is the behaviour which would be expected of an n -cavity structure of the type mentioned when the operating frequency is varied.

7. Application to a practical structure

As an example of the application of the method developed above, consider a structure of the type used in the Berkeley proton accelerator (2). This is developed from the periodic structure formed by the repetition of elements sketched (in somewhat idealized form) in Fig. 2. The protons

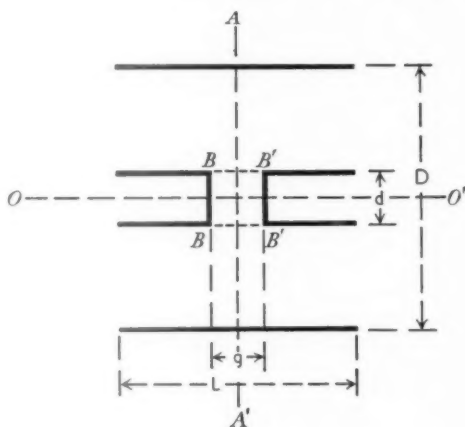


FIG. 2

travel along the axis of symmetry OO' and are accelerated by the electric field in each gap BB' . The length L , which is determined by the relation $L/\lambda = v/c$ (λ the operating wavelength, v the velocity of the protons, c the velocity of light) increases slowly along the structure to allow for the acceleration of the protons; each cavity is designed to be a closed-circuit resonant cavity by making the dimensions g and d vary with L , the outside diameter D being constant. (The resonant mode is one in which the only non-vanishing field components are the cylindrical polar components E_r , E_z , and H_θ .) In the Berkeley accelerator this variation is achieved in the early stages by increasing g/L , keeping d constant, and later by increasing d keeping g/L constant. For simplicity we consider only the former type of variation. (If d is varied it is necessary to take the dividing plane between cavities as the central plane AA' of Fig. 2 in order to have congruent apertures.) If each resonator is accurately tuned, the input-output voltage and current relations are given by (6.3), and if the accelerator is terminated by a short circuit the voltage is zero at all the apertures. The variation of current (and hence of the electromagnetic field strength) will be determined by the transformer action at the junction of adjoining cavities ((4.1) and (4.2)), and it is this variation that we calculate.

The resonant field in the cavity can be calculated approximately by assuming that the longitudinal component of the electric field on the cylinder BB' which forms the continuation of the inner conductor is constant. The field in the space between the coaxial cylinders can be obtained as the sum of a series of coaxial waveguide modes, and the aperture field \mathcal{H} is thus determined. In a similar approximation the corresponding 'odd-magnetic' field of section 6, which determines the aperture field \mathcal{E} , can be calculated assuming zero longitudinal electric field on BB' . The integrals $\int_S \mathcal{E} \cdot \mathcal{H} dS$ and $\int_S \mathcal{E} \cdot \mathcal{H}' dS$, which determine the ratio of currents in adjacent cavities, can be obtained conveniently from these series in virtue of the orthogonality relations of the waveguide modes.

It is convenient to define unit amplitude of the resonant mode in the cavity as that which corresponds to unit value of the mean longitudinal electric field on BB' . With this definition \mathcal{H}_θ is found to have the value

$$-\frac{ik}{Z_0} \left[\sum_{n=1}^{\infty} \frac{\pi \kappa_n}{m_n^2} \frac{J_0(\kappa_n d/2) J_0(\kappa_n D/2)}{\{J_0(\kappa_n d/2)\}^2 - \{J_0(\kappa_n D/2)\}^2} \frac{\sinh(m_n g/2)}{\sinh(m_n L/2)} \mathcal{J}_1(\kappa_n r) - \frac{1}{k^2 \log(D/d)} \frac{\sin(kg/2)}{\sin(kL/2)} \frac{1}{r} \right],$$

where $\kappa_1, \kappa_2, \dots$ are the roots of the equation

$$J_0(\kappa d/2) Y_0(\kappa D/2) - Y_0(\kappa d/2) J_0(\kappa D/2) = 0,$$

$m_n = \sqrt{(\kappa_n^2 - k^2)}$, J_0 and Y_0 are the Bessel functions of first and second kind of order zero, and

$$\mathcal{J}_1(\kappa_n r) = Y_0(\kappa_n D/2) J_1(\kappa_n r) - J_0(\kappa_n D/2) Y_1(\kappa_n r).$$

\mathcal{E}_r is

$$-\frac{1}{2} \left[\sum_{n=1}^{\infty} \frac{\pi \kappa_n}{m_n} \frac{J_0(\kappa_n d/2) J_0(\kappa_n D/2)}{\{J_0(\kappa_n d/2)\}^2 - \{J_0(\kappa_n D/2)\}^2} \frac{\sinh(m_n g/2)}{\sinh(m_n L/2)} \coth(m_n L/2) \mathcal{J}_1(\kappa_n r) - \frac{1}{k \log(D/d)} \frac{\sin(kg/2)}{\sin(kL/2)} \cot(kL/2) \frac{1}{r} \right].$$

The ratio $i'/i = \int_S \mathcal{E}_r \mathcal{H}_\theta dS / \int_S \mathcal{E}_r \mathcal{H}'_\theta dS$ of currents in adjacent sections, in one of which L is replaced by $L' = L + \delta L$ and g by $g' = g + \delta g$, can be found to first order in δL and δg by evaluating the integrals and differentiating with respect to L' and g' . The result can be expressed in the form

$$\delta \log i = (1 + F_L) \delta L / L - (1 + F_g) \delta g / g$$

where

$$F_L = \frac{\sum_{n=1}^{\infty} G_n \{ (m_n L/2) \coth(m_n L/2) - 1 \} + G_0 \{ (kL/2) \cot(kL/2) - 1 \}}{\sum_{n=1}^{\infty} G_n + G_0},$$

$$F_g = \frac{\sum_{n=1}^{\infty} G_n \{ (m_n g/2) \coth(m_n g/2) - 1 \} + G_0 \{ (kg/2) \cot(kg/2) - 1 \}}{\sum_{n=1}^{\infty} G_n + G_0},$$

$$G_n = \frac{1}{m_n^3} \frac{\{J_0(\kappa_n D/2)\}^2}{\{J_0(\kappa_n d/2)\}^2 - \{J_0(\kappa_n D/2)\}^2} \frac{\sinh^2(m_n g/2)}{\sinh^2(m_n L/2)} \coth(m_n L/2),$$

$$G_0 = \frac{1}{2k^3 \log(D/d)} \frac{\sin^2(kg/2)}{\sin^2(kL/2)} \cot(kL/2).$$

F_L and F_g are small compared with unity for the dimensions considered. If they are neglected the equation for i can be integrated and shows that ig/L is the same for all cavities. ig/L is proportional to the local mean value \bar{E} of the longitudinal electric field along the axis of the system, and therefore to the velocity increment per cavity. The design of the Berkeley accelerator was based on the assumption of constancy of this quantity.

Values of F_L and F_g were computed for the range of values

$$0.09 < L/\lambda < 0.18,$$

with the values $D/\lambda = 0.66$, $d/D = 0.124$, used in the Berkeley accelerator. The values of g for resonance were obtained from the empirical formula given by Alvarez (2), the range of g/L being from about 0.2 to 0.3. In this range it was found that F_g was negligible (< 0.005) and that F_L could be accurately represented by the formula $F_L = -0.64(L/D)^2$. With this formula the equation for i shows that \bar{E} is proportional to $e^{-0.32(L/D)^2}$, giving a variation of about 2 per cent over the range of values considered.

A variation of this amount would be completely masked by tuning errors, but the form of dependence on L suggests that the effect might be important at the high velocity end of the accelerator.

REFERENCES

1. J. C. SLATER, *Rev. Mod. Phys.* **20** (1948) 473.
2. L. W. ALVAREZ *et al.*, *Rev. Sci. Inst.* **26** (1955) 111.
3. L. BRILLOUIN, *Wave Propagation in Periodic Structures* (McGraw-Hill, 1946).
4. J. J. WILKINS, A.E.R.E. GPR 1613.

ON A CLASS OF DIFFERENTIAL EQUATION GOVERNING NON-LINEAR VIBRATIONS

By A. W. GILLIES (*Northampton Polytechnic, London*)

[Received 4 October 1956]

SUMMARY

A differential equation of the form

$$f(D)x + g(D)y = F(t)$$

is considered, with y a function of x , expansible in a series in powers of x ; $F(t)$ is a simple periodic function and D is the operator d/dt . The approximate linear equation when x and y are small is assumed to have one nearly sinusoidal normal mode, its other modes being exponentially damped with time constants less than a constant which is $O(1)$.

The method previously applied to particular second- and third-order equations of this form is applied to determine non-resonant and resonant periodic solutions of this equation. The solutions considered are of the form $2b \cos(\omega t + \phi)$ plus terms of higher order, where b and ϕ are either constants satisfying an amplitude equation or approximate to slowly varying functions of the time, satisfying an autonomous system of variational equations. These are of the same form as in the particular cases. Thus, although the equation is of arbitrarily high order, it behaves, as far as the solutions considered in this paper are concerned, like one of the second order. Although the solution of an equation of order n depends in general on n arbitrary constants, in this case, in consequence of the form assumed for the linearized equation, only the two parameters b and ϕ are of importance, apart from a transient stage, the duration of which is of the order of the longest time constant of the linearized equation.

The variational equations are of four main types; in two of these the final state is a periodic solution or a combination oscillation determined by a singular point or limit cycle of the variational equations, in the other two the oscillation amplitude may increase until the variational equations cease to be valid, and the final state cannot be determined by the methods of this paper.

1. Introduction

IN a large class of oscillation phenomena one is concerned with a system which is substantially linear with the exception of one non-linear connexion, which leads to a differential equation of the form

$$f(D)x + g(D)y = F(t), \quad (1)$$

$f(D)$ and $g(D)$ being polynomials in the operator $D \equiv d/dt$, with a non-linear functional relation between the dependent variables x and y . If this relation is analytic then for sufficiently small values of the variables a power series expansion of the form

$$y = c_1 x + c_2 \mu x^2 + c_3 \mu^2 x^3 + \dots \quad (2)$$

in which c_1 may be assumed positive (otherwise y could be replaced by $-y$).

the c_r in general are assumed to be $O(1)$ and μ is a small positive parameter characterizing the magnitude of the non-linear terms. The series will converge in some interval $\mu|x| < R_1$.

In practice the analytic form of the relationship between x and y will often be unknown and will be replaced by a polynomial approximation to an experimentally determined relationship. The interval of convergence will then be replaced by the interval of validity of this polynomial approximation.

In either case it will be assumed, not only that the variables are restricted to an interval in which (2) is valid, but that furthermore the non-linear terms are small, so that in the equations of the first approximation only the first three terms of (2) are used.

With x small enough for non-linear terms to be neglected, and with the forcing term $F(t)$ also omitted, we have the approximate linear equation

$$\{f(D) + c_1 g(D)\}x = 0. \quad (3)$$

The possibility of free oscillation, and of resonance and synchronization phenomena with a periodic forcing term, will arise when this linear equation has an oscillatory normal mode which is at most only slightly damped.

It is therefore assumed that the polynomial

$$G(z) = f(z) + c_1 g(z)$$

can be factorized in the form

$$G(z) = \{(z - \epsilon)^2 + \Omega^2\}h(z).$$

The linear equation (3) then has a normal mode of the form

$$Ae^{\epsilon t} \cos(\Omega t + \phi)$$

with A and ϕ arbitrary constants, and if ϵ is small in comparison with Ω this will approximate to a simple periodic function, the fractional change in amplitude in one period being small in comparison with unity.

By a change of time scale the equation may be normalized to give the period 2π for this normal mode, when we shall have

$$G(z) = \{(z - \epsilon)^2 + 1\}h(z)$$

with ϵ small in comparison with unity.

It will be further assumed that all other modes of (3) are damped with time constant less than T , where T is a constant not greater than $O(1)$; i.e. all the zeros of $h(z)$ are to the left of a line $\operatorname{Re} z = -1/T$ on the complex plane. $|h(i\omega)|$ will then have a minimum value, say $m > 0$, as ω varies from $-\infty$ to $+\infty$.

$|G(i\omega)|$ will then have a minimum value greater than $m/2$ if ω^2 is outside the interval $\frac{1}{2} < \omega^2 < \frac{3}{2}$ around the resonant frequency value 1.

It is also assumed that the degree of $g(z)$ is not greater than that of $G(z)$, i.e. that the order of the original equation (1) is the same as that of the linear equation (3). In other words, regarded as a perturbation problem, it is non-singular.

Finally, the forcing term is taken as a simple periodic function

$$F(t) = 2B \cos \omega t.$$

Thus, on eliminating y we have an equation of the form

$$\{f(D) + c_1 g(D)\}x + g(D) \sum_{n=2}^{\infty} c_n \mu^{n-1} x^n = 2B \cos \omega t$$

$$\text{or} \quad \{(D - \epsilon)^2 + 1\}h(D)x + g(D) \sum_{n=2}^{\infty} c_n \mu^{n-1} x^n = 2B \cos \omega t, \quad (4)$$

in which

I. ϵ and μ are small parameters with $\mu > 0$;

II. $\min |h(i\omega)| = m > 0$ for $-\infty < \omega < +\infty$;

from which follows

$$\min \{|(i\omega - \epsilon)^2 + 1\}h(i\omega)| > \frac{1}{2}m \quad \text{for } 0 < \omega^2 < \frac{1}{2} \text{ and } \frac{3}{2} < \omega^2;$$

III. the zeros of $h(z)$ are all to the left of $\operatorname{re} z = -1/T$ with T a constant not greater than $O(1)$;

IV. the degree of $g(z)$ is not greater than that of

$$f(z) + c_1 g(z) = \{(z - \epsilon)^2 + 1\}h(z).$$

It is the purpose of the present paper to consider the periodic solutions of (4) subject to conditions I to IV with particular reference to the resonance case when ω is near to 1, i.e. $\omega - 1 = O(\epsilon)$, B is $O(\epsilon)$, and the small parameters are connected by $|\epsilon| = \mu^2$. These latter assumptions will be brought in when required.

In previous papers, particular second order (1) and third order (2) equations of this general type have been discussed. The same procedure is applied to (4) and results of the same form are obtained. Although the equation is of arbitrarily high order, and the solution is dependent on a corresponding number of arbitrary constants, it appears that, owing to the form assumed for the linearized equation, after a transient interval it behaves (as far as the solutions considered in the paper are concerned), roughly like an equation of the second order. These solutions are of the type $2b \cos(\omega t + \phi)$ plus terms of higher order, where b and ϕ are either constants that satisfy an amplitude equation [(17) and (18) below] or slowly varying functions of the time that satisfy an autonomous system of variational equations [(37) and (38) below]. These equations, as well

as the resonance curves, are of the same forms as were obtained in the particular cases.

2. The non-resonant solution

We seek to determine a periodic solution of (4) in powers of B . Set

$$x = x^{(1)}B + x^{(2)}B^2 + x^{(3)}B^3 + \dots \quad (5)$$

Substitute in (4) and determine the $x^{(n)}$ in succession as periodic functions of t by equating coefficients of like powers of B .

Writing
$$\zeta(z) = \frac{f(z) + c_1 g(z)}{g(z)}$$

(4) and

$$\zeta_r = \zeta(ri\omega),$$

the first few $x^{(n)}$ are with

$$x^{(1)} = x_1^{(1)} + x_{-1}^{(1)},$$

$$x_1^{(1)} = \frac{1}{f(i\omega) + c_1 g(i\omega)} e^{i\omega t},$$

$$x_{-1}^{(1)} = \text{conjugate of } x_1^{(1)};$$

$$x^{(2)} = x_2^{(2)} + x_0^{(2)} + x_{-2}^{(2)}, \quad (6)$$

with

$$x_2^{(2)} = -\frac{c_2 \mu x_1^{(1)2}}{\zeta_2},$$

$$x_0^{(2)} = -\frac{2c_2 \mu x_1^{(1)} x_{-1}^{(1)}}{\zeta_0},$$

$$x_{-2}^{(2)} = \text{conj. of } x_2^{(2)};$$

$$x^{(3)} = x_3^{(3)} + x_1^{(3)} + x_{-1}^{(3)} + x_{-3}^{(3)},$$

with

$$x_3^{(3)} = -\frac{2c_2 \mu x_1^{(1)} x_2^{(2)} + c_3 \mu^2 x_1^{(1)3}}{\zeta_3} = \left(\frac{c_2^2}{\zeta_2} - c_3 \right) \frac{\mu^2 x_1^{(1)3}}{\zeta_3}, \quad (7)$$

$$x_1^{(3)} = -\frac{2c_2 \mu (x_1^{(1)} x_0^{(2)} + x_{-1}^{(1)} x_2^{(2)}) + 3c_3 \mu^2 x_1^{(1)2} x_{-1}^{(1)}}{\zeta_1}$$

$$= \left(\frac{4c_2^2}{\zeta_0} + \frac{2c_2^2}{\zeta_2} - 3c_3 \right) \frac{\mu^2 x_1^{(1)2} x_{-1}^{(1)}}{\zeta_1},$$

and $x_{-3}^{(3)}, x_{-1}^{(3)}$ the conjugates of these. The general form is

$$x^{(n)} = \sum_r x_r^{(n)},$$

the summation being over alternate integers from $-n$ to $+n$. The term $x_r^{(n)}$ contains the exponential factor $e^{ri\omega t}$ and combines with its conjugate $x_{-r}^{(n)}$ to form a real term of r th harmonic frequency.

With these values for the $x^{(n)}$ the series (5) gives a formal solution of the differential equation which is of period 2π .

It may be noted that $x^{(n)}$ contains μ^{n-1} as a factor, so that the series may be regarded as a series in powers of μ .

3. Convergence of the non-resonant solution

The absolute and uniform convergence with respect to t of the foregoing solution may be established by the aid of the equation

$$\xi = N\eta - c\mu_1\eta^2 - c\mu_1^2\eta^3 - \dots$$

in which c is a positive constant greater than $|c_n|$ ($n = 2, 3, \dots$) and N is a positive constant to be chosen later.†

This is equivalent to

$$\xi = (N+c)\eta - \frac{c\eta}{1-\mu_1\eta}$$

which defines η as a two-valued function of ξ . The branch which vanishes for $\xi = 0$ may be expanded in powers of ξ in a circle extending to the branch point nearer to the origin. This is at $\xi = R$, $\eta = S$, where

$$R = \frac{1}{\mu_1} [N + 2c - 2(Nc + c^2)^{\frac{1}{2}}],$$

$$S = \frac{1}{\mu_1} \left[1 + \left(\frac{c}{N+c} \right)^{\frac{1}{2}} \right]. \quad (8)$$

If this series for η in powers of ξ is obtained by reversion of the former series it will be seen to be a majorante of (5) provided

$$\mu < \mu_1, \quad (9)$$

$$\frac{2B}{|f(i\omega) + c_1 g(i\omega)|} < \frac{\xi}{N} < \frac{R}{N}, \quad (10)$$

$$|\xi_r| > N \quad (r = 0, \pm 1, \pm 2, \dots). \quad (11)$$

In fact, if the series is written

$$\eta = \eta^{(1)}\xi + \eta^{(2)}\xi^2 + \eta^{(3)}\xi^3 + \dots \quad (12)$$

and (9), (10), (11) are satisfied, we have

$$|x^{(n)}|B^n \leq \sum |x_r^{(n)}|B^n < \eta^{(n)}\xi^n.$$

The series (5) will then converge absolutely and uniformly with respect to t , provided

$$B < B_1 = \frac{R|f(i\omega) + c_1 g(i\omega)|}{N} \quad (13)$$

for any $\mu < \mu_1$; and it will then provide a periodic solution of the differential equation (4).

† In particular cases a wider interval of convergence may be obtained using

$$\xi = N\eta - |c_2|\mu_1\eta^2 - |c_3|\mu_1^2\eta^3 - \dots,$$

but this is less convenient for the general discussion as it cannot be expressed in finite form.

If ϵ is independent of μ , then B_1 may be chosen arbitrarily, and μ_1 determined by (13). We then have convergence in any specified interval of B , provided μ is sufficiently small, i.e. provided the non-linear terms are sufficiently small.

If, however, ϵ is small this may require that μ is smaller than $O(\epsilon^2)$ if ω is unrestricted. This arises from the fact that N must be taken $O(\epsilon)$ to satisfy (10) if any $r\omega$ is near to 1. [Note that condition IV ensures that $\zeta(z)$ does not tend to zero at infinity.] That is, if resonance occurs the series solution (5) will only have an interval of convergence of order ϵ^2/μ .

4. The resonant solution

When an $r\omega$ is near to 1 the solution of the previous sections ceases to be useful when ϵ is small. The failure arises through the occurrence of a small value, of $O(\epsilon)$, for one of the ζ_r divisors in the determination of the $x^{(n)}$. If this particular divisor can be avoided, it may again be possible to choose N independently of ϵ and so secure a useful interval of convergence. This will be done for the case of fundamental resonance when ϵ is near to 1 and $\zeta_{\pm 1}$ are $O(\epsilon)$.

We suppose that a periodic solution exists whose fundamental is $2b \cos(\omega t + \phi)$ and set

$$x = x^{(1)}b + x^{(2)}b^2 + x^{(3)}b^3 + \dots \quad (14)$$

with $x^{(1)} = x_1^{(1)} + x_{-1}^{(1)}$,

where $x_1^{(1)} = e^{i(\omega t + \phi)}$, $x_{-1}^{(1)} = e^{-i(\omega t + \phi)}$.

Thus $x^{(1)}b = 2b \cos(\omega t + \phi)$ is the complete fundamental.

Substitute (14) into (4) and choose the $x^{(n)}$ so as to annul all terms containing b^n other than those of fundamental frequency. The $x^{(n)}$ are then given in terms of $x_1^{(1)}$ and $x_{-1}^{(1)}$ by the same expressions as before with the omission of $x_1^{(n)}$ and $x_{-1}^{(n)}$ when n is odd, and of terms arising from these in the higher orders.

The terms which were previously cancelled by the $x_{\pm 1}^{(n)}$ now remain to form an equation containing only terms with $e^{i\omega t}$ and their conjugates with $e^{-i\omega t}$. This equation will be satisfied if the equation formed by the terms in $e^{i\omega t}$ only is satisfied; then $x^{(2)}$ will be given by (6) and the following equations, and $x^{(3)}$ will now contain only $x_3^{(3)}$ as given by (7) and $x_{-3}^{(3)}$. The terms in $e^{i\omega t}$ which remain are

$$\{f(D) + c_1 g(D)\} x_1^{(1)} b - g(D) \left\{ \left(\frac{4c_2^2}{\zeta_0} + \frac{2c_2^2}{\zeta_2} - 3c_3 \right) \mu^2 x_1^{(1)2} x_{-1}^{(1)} b^3 + \dots \right\} = B e^{i\omega t} \quad (15)$$

$$\text{or} \quad \{f(i\omega) + c_1 g(i\omega)\} b - g(i\omega) \left\{ \left(\frac{4c_2^2}{\zeta_0} + \frac{2c_2^2}{\zeta_2} - 3c_3 \right) \mu^2 b^3 + \dots \right\} = B e^{-i\phi}. \quad (16)$$

Equation (16), equivalent to two real equations, is the amplitude equation, and if values of b and ϕ exist which satisfy this equation and for which the series (14) converges, then each such pair will determine a periodic solution of the differential equation (4).

By this procedure $\zeta_{\pm 1}$ do not occur as divisors, so N has to satisfy (10) only for $r = 0, \pm 2, \pm 3, \dots$ and not for $r = 1$. If ω is near to 1 and no $r\omega$ ($r = 0, \pm 2, \pm 3, \dots$) falls within $\frac{1}{2} < r^2\omega^2 < \frac{3}{2}$, then by II it is possible to choose N independently of ϵ ; (12) is now a majorante of (14) provided (9), (11) with $r \neq 1$, are satisfied, and instead of (10)

$$b < R/2N$$

with R still given by (8).

We can now choose b_1 , then μ_1 , so that $b_1 < \frac{1}{2}R/N$ and we have that (14) converges absolutely and uniformly with respect to t for $b < b_1$ and $\mu < \mu_1$.

Furthermore, the sum of the moduli of the terms containing b^n in the amplitude equation is less than $N\eta^{(n)}\xi^n$ so that (12) multiplied by N is a majorante of the series in the amplitude equation (16). This therefore converges in the same interval.

The determination of periodic solutions of period $2\pi/\omega$ and amplitude less than b_1 is now reduced to the determination of solutions of the amplitude equation.

It may be noted that $x^{(n)}$ in (14) contains the factor μ^{n-1} , so that (14) may be regarded as a series in powers of μ . Likewise the amplitude equation has the general form

$$\frac{Be^{-i\phi}}{g(i\omega)} = \sum_{n=0}^{\infty} A_n \mu^{2n} b^{2n+1},$$

where $A_0 = \zeta_1$ and A_n ($n \geq 1$) is a function of the coefficients c_2, \dots, c_{2n+1} and of the $\zeta_{\pm r}$ for $r = 0, \pm 2, \pm 3, \dots, \pm 2n$. Thus this equation also may be regarded as a series in powers of μ . In fact this viewpoint is appropriate in considering the orders of magnitude of the terms, once we have fixed the interval of convergence $b < b_1$.

5. The solution of the amplitude equation

We know already that if B is sufficiently small and $\zeta_1 = \zeta(i\omega)$ is not zero, the amplitude equation has one solution, given by the non-resonant solution, for which $b \rightarrow 0$ with B . The non-resonant solution could in fact be recovered by reversion of (16).

To consider further the solution of the amplitude equation, we must make some assumptions regarding the magnitude of its terms. We therefore introduce the following:

- V. the small parameters are related by $|\epsilon| = \mu^2$;
 VI. the amplitude of the forcing term is $O(\epsilon)$ with $B = |\epsilon|E$;
 VII. the frequency ω differs from 1 by $O(\epsilon)$ with $\omega - 1 = |\epsilon|\sigma$.

With these three assumptions we have

$$f(i) + c_1 g(i) = -2ih(i)\epsilon + O(\epsilon^2) = \epsilon A e^{i\alpha} + O(\epsilon^2)$$

with

$$A = 2|h(i)|, \quad \alpha = \arg h(i) - \frac{1}{2}\pi;$$

$$f'(i) + c_1 g'(i) = 2ih(i) + O(\epsilon) = -A e^{i\alpha} + O(\epsilon).$$

So

$$f(i\omega) + c_1 g(i\omega) = A e^{i\alpha} (\delta - i\sigma) |\epsilon| + O(\epsilon^2)$$

where

$$\delta = \frac{\epsilon}{|\epsilon|} = \pm 1.$$

The factor occurring in the coefficient of b^3 in (16) is

$$\frac{4c_2^2}{\zeta_0^2} + \frac{2c_2^2}{\zeta_2^2} - 3c_3 = \frac{4c_2^2}{\zeta_0^2} + \frac{2c_2^2}{\zeta(2i)} - 3c_3 + O(\epsilon) = \lambda' + i\lambda'' + O(\epsilon),$$

say, and

$$g(i\omega) = g(i) + O(\epsilon).$$

The amplitude equation now becomes

$$A e^{i\alpha} (\delta - i\sigma) b - g(i) (\lambda' + i\lambda'') b^3 + O(\mu^2) = E e^{-i\phi}.$$

Omitting the $O(\mu^2)$ we have the equation of the first approximation, which may be rearranged as

$$(\delta - i\sigma) b - \frac{g(i)(\lambda' + i\lambda'')}{A e^{i\alpha}} b^3 = \frac{E}{A} e^{-i(\phi + \alpha)}.$$

Let

$$\frac{g(i)(\lambda' + i\lambda'')}{A e^{i\alpha}} = \Lambda e^{i\lambda}$$

and separate real and imaginary parts. Then

$$\delta b - \Lambda \cos \lambda b^3 = \frac{E}{A} \cos(\phi + \alpha), \quad (17)$$

$$\sigma b + \Lambda \sin \lambda b^3 = \frac{E}{A} \sin(\phi + \alpha), \quad (18)$$

which gives the real equations of the first approximation for b and ϕ .

When the linear equation has an increasing oscillatory mode ϵ is positive and $\delta = +1$. In this case (17), with $E = 0$, gives for the free oscillation

$$b = 0 \quad \text{or} \quad b^2 = \frac{1}{\Lambda \cos \lambda},$$

whilst (18) gives for the second value of b

$$\sigma = -\Lambda \sin \lambda b^2 = -\tan \lambda.$$

For a free oscillation to exist, $\cos \lambda$ must be positive and not small of $O(\mu)$

for the value given for b to fall within the interval of convergence $b < b_1$. For this to be the case we require that

VIII. $-\frac{1}{2}\pi + \Delta < \lambda < \frac{1}{2}\pi - \Delta$ with $\epsilon > 0$.

When this condition is satisfied, let the free oscillation amplitude be a ; (17), (18) may then be rearranged as

$$\frac{b}{a} \left(1 - \frac{b^2}{a^2} \right) = \frac{E \cos(\phi + \alpha)}{Aa},$$

$$\frac{b}{a} \left(\sigma + \nu \frac{b^2}{a^2} \right) = \frac{E \sin(\phi + \alpha)}{Aa},$$

where

$$\nu = \tan \lambda.$$

If we write $E/(Aa) = F$, $\phi + \alpha = \phi'$, $b/a = b'$, these equations become

$$b'(1 - b'^2) = F \cos \phi', \quad (19)$$

$$b'(\sigma + \nu b'^2) = F \sin \phi'. \quad (20)$$

These are of the same standard form obtained in the particular cases (1), (2). The solution of these equations has been considered previously and typical resonance curves (graphs of b^2 against σ for fixed F), are given in (1), Figs. 1 and 2, for $\nu = 1$ and $\nu = 2$, respectively. For $\nu = 0$ they are the symmetrical van der Pol resonance curves. The curves for $\nu < 0$ are obtained by reflecting those for $\nu > 0$ in the b^2 axis.

These resonance curves are therefore typical of a non-linear oscillatory system having a nearly sinusoidal free oscillation.

We may use implicit function theory to show that the complete equations for b , ϕ have solutions when $|\epsilon| = \mu^2$ is sufficiently small, differing at most by $O(\epsilon)$ from the solutions of the equations of the first approximation.

6. Differential equations for the amplitude and phase

Denote the formal solution (14) by

$$\begin{aligned} K(b, \phi, t) &= x^{(1)}b + x^{(2)}b^2 + x^{(3)}b^3 + \dots \\ &= 2b \cos(\omega t + \phi) + O(\mu). \end{aligned} \quad (21)$$

When $K(b, \phi, t)$ is substituted for x in (4) the residual terms form the amplitude equation.

Write the differential equation (4) in the form

$$(D^2 + 1)h(D)x - (2\epsilon D - \epsilon^2)h(D)x + g(D) \sum_{n=2} c_n \mu^{n-1} x^n - 2B \cos \omega t = 0 \quad (22)$$

or

$$(D^2 + 1)h(D)x + \mu S(x) - 2B \cos \omega t = 0,$$

in which $\mu S(x)$ denotes the complete group of small terms, and is actually a function of x and its differential coefficients, i.e.

$$\mu S(x) = -(2\mu^2 D - \mu^4)h(D)x + g(D) \sum_{n=2} c_n \mu^{n-1} x^n.$$

The residual terms on the left when K is substituted for x are

$$\left[\{f(i\omega) + c_1 g(i\omega)\}b - g(i\omega) \left(\frac{4c_2^2}{\zeta_0} + \frac{2c_2^2}{\zeta_2} - 3c_3 \right) \mu^2 b^3 + \dots \right] e^{i\theta} - B e^{i\omega t} + \text{conj.}^i$$

in which $\theta = \omega t + \phi$. Let these terms be denoted by $2 \frac{dH(b, \phi, t)}{dt}$, i.e.

$$\begin{aligned} H(b, \phi, t) &= \frac{1}{2i\omega} [e^{i\theta} \{f(i\omega) + c_1 g(i\omega)\}b + \dots] - B e^{i\omega t} + \text{conj.} \\ &= \frac{\mu^2}{2i} [e^{i\theta} \{A e^{i\alpha} (\delta - i\sigma)b - g(i)(\lambda' + i\lambda'')b^3\} - E e^{i\omega t} + O(\mu^2)] + \text{conj.} \end{aligned}$$

If b, ϕ are variables along with t we have

$$\left(\frac{\partial^2}{\partial t^2} + 1 \right) h \left(\frac{\partial}{\partial t} \right) K + \mu S(K) - 2B \cos \omega t = 2 \frac{\partial H}{\partial t} \quad (23)$$

$$\mu S(K) = - \left(2\mu^2 \frac{\partial}{\partial t} - \mu^4 \right) h \left(\frac{\partial}{\partial t} \right) K + g \left(\frac{\partial}{\partial t} \right) \sum_{n=2} c_n \mu^{n-1} K^n.$$

The differential equation (22) is now replaced by the system of two equations

$$(D^2 + 1)u + \mu S(x) - 2B \cos \omega t = 0, \quad (24)$$

$$u = h(D)x, \quad (25)$$

the formal solution of which is

$$\begin{aligned} x &= K(b, \phi, t), \\ u &= h(D)K(b, \phi, t) \\ &= K_1(b, \phi, t), \end{aligned}$$

say, with b, ϕ constants satisfying the amplitude equation, i.e. for which H and dH/dt both vanish. Now take b, ϕ as new variables by using

$$u = K_1(b, \phi, t), \quad (26)$$

$$\frac{du}{dt} = \frac{\partial K_1}{\partial t} - H. \quad (27)$$

$$\text{These require that } \frac{\partial K_1}{\partial b} \frac{db}{dt} + \frac{\partial K_1}{\partial \phi} \frac{d\phi}{dt} + H = 0. \quad (28)$$

Substituting into (24), using (23) and (25), we obtain

$$\frac{\partial}{\partial b} \left(\frac{\partial K_1}{\partial t} - H \right) \frac{db}{dt} + \frac{\partial}{\partial \phi} \left(\frac{\partial K_1}{\partial t} - H \right) \frac{d\phi}{dt} + \mu \{S(x) - S(K)\} + \frac{\partial H}{\partial t} = 0. \quad (29)$$

Solving (28) and (29) for db/dt and $d\phi/dt$,

$$\frac{db}{dt} = \frac{1}{\Delta_1} \left\{ L \frac{\partial K_1}{\partial \phi} - H \frac{\partial}{\partial \phi} \left(\frac{\partial K_1}{\partial t} - H \right) \right\}, \quad (30)$$

$$\frac{d\phi}{dt} = \frac{1}{\Delta_1} \left\{ -L \frac{\partial K_1}{\partial b} + H \frac{\partial}{\partial b} \left(\frac{\partial K_1}{\partial t} - H \right) \right\}, \quad (31)$$

where

$$L = \mu \{S(x) - S(K)\} + \frac{\partial H}{\partial t}$$

and

$$\Delta_1 = \frac{\partial K_1}{\partial b} \frac{\partial}{\partial \phi} \left(\frac{\partial K_1}{\partial t} - H \right) - \frac{\partial K_1}{\partial \phi} \frac{\partial}{\partial b} \left(\frac{\partial K_1}{\partial t} - H \right).$$

Instead of the original equation we now have the system of three equations (30), (31) together with

$$h(D)x = K_1(b, \phi, t) \quad (32)$$

in the three unknowns b, ϕ, x . These are satisfied if b, ϕ have constant values satisfying the amplitude equation, and $x = K(b, \phi, t)$, i.e. by the periodic solutions of the original equation.

The right-hand members of (30), (31) may be expanded in powers of μ . Evaluating only the terms of lowest index we obtain

$$\begin{aligned} \frac{db}{dt} = & \frac{\mu}{A} \{S(x) - S(K)\} \{\cos(\theta + \alpha) + O(\mu)\} + \\ & + \mu^2 \left\{ \delta b - \Lambda \cos \lambda b^3 - \frac{E}{A} \cos(\phi + \alpha) \right\} + O(\mu^3), \end{aligned} \quad (33)$$

$$\begin{aligned} b \frac{d\phi}{dt} = & -\frac{\mu}{A} \{S(x) - S(K)\} \{\sin(\theta + \alpha) + O(\mu)\} - \\ & - \mu^2 \left\{ \sigma b + \Lambda \sin \lambda b^3 - \frac{E}{A} \sin(\phi + \alpha) \right\} + O(\mu^3). \end{aligned} \quad (34)$$

Under VIII, when a free oscillation is possible, these become with the notation of (19), (20)

$$\begin{aligned} \frac{db'}{dt} = & \frac{\mu}{Aa} \{S(x) - S(K)\} \{\cos(\omega t + \phi') + O(\mu)\} + \\ & + \mu^2 \{b'(1 - b'^2) - F \cos \phi'\} + O(\mu^3), \\ b' \frac{d\phi'}{dt} = & -\frac{\mu}{Aa} \{S(x) - S(K)\} \{\sin(\omega t + \phi') + O(\mu)\} - \\ & - \mu^2 \{b'(\sigma + \nu b'^2) - F \sin \phi'\} + O(\mu^3), \end{aligned}$$

which are of the form discussed in (2).

7. Qualitative character of the solutions

If the differential equation (4) is of order n^* , a solution will be determined if values are given for $x, Dx, \dots, D^{n^*-1}x$ at $t = 0$. Such a solution may be represented by a trajectory Γ in a space \mathcal{S} in which $x, Dx, \dots, D^{n^*-1}x$ are cartesian coordinates; the trajectory starting from a point P_0 whose co-ordinates are the initial values.

The space \mathcal{S} is transformed into \mathcal{S}' by the non-singular transformation

$$\begin{aligned}x &= x \\D^{n^*-3}x &= D^{n^*-3}x \\u &= h(D)x \\du/dt &= h(D)Dx,\end{aligned}$$

$h(z)$ being of degree n^*-2 ; Γ will transform into Γ' .

A further transformation to \mathcal{S}'' is made by (26), (27), polar coordinates b, ϕ replacing u and du/dt , the other coordinates being unchanged.

The latter transformation is only meaningful provided $b < b_1$. To the interior of the circle $b < b_1$ in the b, ϕ plane of \mathcal{S}'' corresponds a region $u^2 + (du/dt)^2 < Ab_1 - O(\mu)$ approximating to a circle in the $u, du/dt$ plane of \mathcal{S}' , and this independently of t since the series in (26), (27) converge uniformly with respect to t .

We can therefore certainly define a sphere \mathcal{R} with centre at the origin of \mathcal{S} which transforms into \mathcal{R}' in \mathcal{S}' and \mathcal{R}'' in \mathcal{S}'' such that \mathcal{R}'' is within the cylinder $b < b_1$ for all $t > 0$. Any point P in \mathcal{R} transforms to P' in \mathcal{R}' and P'' in \mathcal{R}'' . The determinant of the transformation from $(u, du/dt)$ to (b, ϕ) is Δ_1 which vanishes only when $b = 0$ and ϕ is indeterminate.

To a trajectory Γ starting at P_0 in \mathcal{R} correspond trajectories Γ' starting at P'_0 in \mathcal{R}' , and Γ'' starting at P''_0 in \mathcal{R}'' . These trajectories may be followed for increasing t , either for all $t > 0$, or until some value of t for which they approach the boundaries of their respective regions when the transformation from \mathcal{R}' to \mathcal{R}'' becomes meaningless.

Since the initial values are bounded, db/dt and $b(d\phi/dt)$ are not greater than $O(\mu)$, so that if P_0 is not near to the boundary the trajectories will remain within their respective regions for an interval of time T_1 which is $O(1/\mu)$. Now by (32)

$$h(D)x = K_1(b, \phi, t) = h\left(\frac{\partial}{\partial t}\right)K(b, \phi, t),$$

$$\text{i.e. } h(D)\{x - K(b, \phi, t)\} = \left\{h\left(\frac{\partial}{\partial t}\right) - h\left(\frac{d}{dt}\right)\right\}K(b, \phi, t) = K_2, \quad (35)$$

where K_2 is a sum of a finite number of terms containing partial differential coefficients of K multiplied by differential coefficients of b and ϕ with respect to t , which will be $O(\mu)$.

Let λ_j be a root of $h(z) = 0$ of multiplicity q_j . Then (35) is equivalent to the integral equation

$$x - K(t) = \sum_j \sum_{q=0}^{q_j-1} \left[C_{qj} t^q e^{\lambda_j t} + C'_{qj} \int_0^t K_2(t-\tau) \frac{\tau^{q-1}}{(q-1)!} e^{\lambda_j \tau} d\tau \right] \quad (36)$$

in which the C'_{qj} are the constants of the partial fraction expansion

$$\frac{1}{h(z)} = \sum_j \sum_{q=0}^{q_j-1} \frac{C'_{qj}}{(z-\lambda_j)^q}$$

and the C_{qj} are constants chosen so that the initial conditions at P_0'' are satisfied. Thus C_{qj} and C'_{qj} are $O(1)$. The summation \sum_j is over the distinct roots of $h(z) = 0$. Let the real part of λ_j be λ'_j , then by III

$$\lambda'_j < -1/T.$$

Any term of the first type in (36) therefore satisfies

$$|C_{qj} t^q e^{\lambda_j t}| < |C_{qj}| t^q e^{-t/T}.$$

If $q = 0$ such a term steadily decreases, and is reduced to $O(\mu)$ in a time which is $O\{T \log(1/\mu)\}$, well within T_1 . This may be regarded as the normal case, in which $h(z)$ has no multiple zeros. In the exceptional case when $h(z)$ has multiple zeros, q will have non-zero values, certainly not exceeding (n^*-2) , in some terms of (36). Such a term is initially zero, rises to a maximum value (omitting the C_{qj}) of $(qT)^q e^{-q}$ at $t = qT$, then decreases, tending to zero as $t \rightarrow \infty$. The maximum value, regarded as a function of q , is 1 at $q = 0$, decreases to a minimum $e^{-1/T}$ at $q = 1/T$, and thereafter increases without limit with further increase in q . Clearly some restriction on q or T is necessary if such terms are to be reduced to $O(\mu)$ in a time of $O\{T \log(1/\mu)\}$. Such a condition would be $q_j T < C$ for each q_j , with C a constant which is $O(1)$. That is, the time constant of normal modes of the linear equation associated with multiple zeros of the characteristic equation must not exceed a limit which is smaller the higher the multiplicity of the zero.

If a condition of this sort is satisfied, all terms of the first sort in (36) will be reduced to $O(\mu)$ in a time which is $O\{T \log(1/\mu)\}$, and since there is only a finite number of them independent of μ , the same will be true of their sum.

Considering now terms of the second type, we shall have $|K_2(t-\tau)| < m_1$, where m_1 is a constant of $O(\mu)$, at least for a time T_1 . For $t < T_1$ we have

$$\begin{aligned} \left| \int_0^t K_2(t-\tau) \frac{\tau^{q-1}}{(q-1)!} e^{\lambda_j \tau} d\tau \right| &< \int_0^t m_1 \frac{\tau^{q-1}}{(q-1)!} e^{-\tau/T} d\tau \\ &< \frac{m_1 T^q}{(q-1)!} \int_0^\infty \left(\frac{\tau}{T}\right)^{q-1} e^{-\tau/T} d\left(\frac{\tau}{T}\right) \\ &< m_1 T^q, \end{aligned}$$

which is $O(\mu)$ so long as m_1 is so. Again, since there is only a fixed number of such terms, their sum will be at most $O(\mu)$.

We have, therefore, the result that for any initial point P_0'' not near to the boundary of \mathcal{R}'' , the difference $x-K$ is reduced to $O(\mu)$ in a time $O\{T \log(1/\mu)\}$. In a further interval of the same order $S(x)-S(K)$ will be $O(\mu)$, therefore db/dt and $b(d\phi/dt)$, and therefore also K_2 , will be $O(\mu^2)$, i.e. m_1 may be taken $O(\mu^2)$, and the difference $x-K$ will be reduced to $O(\mu^2)$ and will thereafter remain not greater than $O(\mu^2)$.

The values of db/dt and $b(d\phi/dt)$ will then be given within $O(\mu^3)$ by the autonomous system

$$\frac{db}{dt} = \mu^2 \left\{ \delta b - \Lambda \cos \lambda b^3 - \frac{E}{A} \cos(\phi + \alpha) \right\}, \quad (37)$$

$$b \frac{d\phi}{dt} = -\mu^2 \left\{ \sigma b + \Lambda \sin \lambda b^3 - \frac{E}{A} \sin(\phi + \alpha) \right\}, \quad (38)$$

obtained from (33), (34) by omitting terms which are $O(\mu^3)$. Consequently in a further interval of $O(1/\mu)$ the projection of Γ'' on the (b, ϕ) plane will not differ by more than $O(\mu^2)$ from the trajectory Γ_1'' of the autonomous system which starts from the same point at the beginning of this interval.

Three cases then arise:

(i) If Γ_1'' approaches a stable singular point, x approaches the periodic solution, the amplitude and phase of which are given in the first approximation by the b and ϕ of the singular point.

(ii) If Γ_1'' winds on to a stable limit cycle, then at the end of the interval the projection of the representative point of Γ'' , which may deviate by $O(\mu^2)$ from the corresponding point of Γ_1'' , will coincide with a point of another trajectory Γ_2'' of the autonomous system. We may in this way obtain a succession of arcs $\Gamma_1'', \Gamma_2'', \Gamma_3'', \dots$ of the autonomous system, such that the projection of Γ'' follows each in turn within $O(\mu^2)$ for a time $O(1/\mu)$. Clearly, unless the initial point was near to a separatrix, all these arcs wind on to the same limit cycle, and ultimately the projection of Γ'' will remain within $O(\mu^2)$ of this limit cycle.

The solution in this case is an almost periodic oscillation having the character of a combination oscillation. The period of the cycle of variation of amplitude and phase will be $O(1/\mu^2)$ and it will be given with a fractional error $O(\mu)$ by the period of the limit cycle of the autonomous system.

(iii) The trajectory Γ_1'' may pass outside the circle $b < b_1$. In this case again the projection of Γ'' will follow a sequence of arcs of trajectories of the autonomous system, which, if the starting-point is not near to a separatrix, will all pass outside the circle. The solution in this case is an oscillation of increasing amplitude which ultimately passes outside the validity of the equations of the first approximation.

8. Differential equation of the second order

In the foregoing section it is assumed implicitly that $n^* > 2$. If $n^* = 2$, then $h(D)$ in (4) is a constant which may be normalized to 1. In this case $x = u$ and $S(x) = S(K)$ throughout. Then (33), (34) are obtained without the terms in $(S(x) - S(K))$, and form a system in b, ϕ only, which is altogether equivalent to the original differential equation. This case is discussed in (1).

9. Significance of the order of the differential equation

The argument in section 7 depends on the fact that terms of $O(\mu^2)$ and higher order are small in comparison with terms of $O(\mu)$, and the latter small in comparison with unity if μ is sufficiently small. At several points it is assumed that the sum of a finite number of terms is $O(\mu)$ if the number of terms is independent of μ and each term is not greater than $O(\mu)$. The number of terms in such cases increases with n^* , the order of the differential equation, and clearly if n^* is large the restriction on μ is more stringent than if n^* is small, in order to ensure that such sums remain small in comparison with unity.

It appears then that small non-linear terms may be more significant in equations of higher order, and one might formulate the rough criterion that $n^*\mu$ must remain a first-order small quantity.

10. The variational equations

The variational equations (37), (38) fall under four main types, according as to whether δ is $+1$ or -1 , i.e. ϵ positive or negative, and whether $\cos \lambda$ is positive or negative. There are two further subsidiary cases if $\cos \lambda$ is so small that the term $\Lambda \cos \lambda b^3$ is of higher order than the other terms retained in the variational equations.

With suitable changes of scale and origin of ϕ these may be reduced to the forms listed below:

1. $\epsilon > 0$ and $\cos \lambda > 0$,

$$\dot{b} = b(1 - b^2) - F \cos \phi,$$

$$b\dot{\phi} = F \sin \phi - b(\sigma + \nu b^2).$$

This is the form obtained under VIII corresponding to (19), (20). It includes the symmetrical ($\nu = 0$) and unsymmetrical ($\nu > 0$) van der Pol equation, (1), and the usual case of the phase shift oscillator ($\nu < 0$), (2).

2. $\epsilon < 0$ and $\cos \lambda > 0$,

$$\dot{b} = -b(1 + b^2) - F \cos \phi,$$

$$b\dot{\phi} = F \sin \phi - b(\sigma + \nu b^2).$$

This case is typified by a slightly damped oscillator with non-linear restoring force.

3. $\epsilon > 0$ and $\cos \lambda < 0$,

$$\dot{b} = b(1+b^2) - F \cos \phi,$$

$$b\dot{\phi} = F \sin \phi - b(\sigma + \nu b^2).$$

4. $\epsilon < 0$ and $\cos \lambda < 0$,

$$\dot{b} = -b(1-b^2) - F \cos \phi,$$

$$b\dot{\phi} = F \sin \phi - b(\sigma + \nu b^2).$$

The trajectories of 3 and 4 are essentially the same as 2 and 1 respectively, but described in the opposite sense as t increases.

The subsidiary cases are

5. $\epsilon > 0$,

$$\dot{b} = b - F \cos \phi,$$

$$b\dot{\phi} = F \sin \phi - b(\sigma + \nu b^2).$$

6. $\epsilon < 0$,

$$\dot{b} = -b - F \cos \phi,$$

$$b\dot{\phi} = F \sin \phi - b(\sigma + \nu b^2).$$

Of these 5 is essentially the same as 3, and 6 is essentially the same as 2. Case 1 is discussed in some detail in (1) and cases 2-5 are discussed in outline in (2), the nature of the singular points and form of the resonance curves being indicated.

In cases 1, 2, and 6 the equations of the first approximation determine the final state of the system for sufficiently small values of μ . In the other three cases trajectories run outwards for large values of b and from suitable initial conditions an oscillation will grow beyond the limits of validity of the equations of the first approximation.

11. Symbolic derivation of the variational equations

The periodic solutions are determined by the amplitude equation. If we suppose that an existing periodic solution is slightly disturbed, so that the subsequent motion may be regarded as an oscillation of slowly varying amplitude and phase, \dot{b} and $b\dot{\phi}$ being $O(\mu^2)$, then we may neglect higher differential coefficients altogether, and take account of the variation of b and ϕ in the first-order terms only. The amplitude equation (15) may then be written

$$G(D)be^{i(\omega t + \phi)} - g(i)(\lambda' + i\lambda'')\mu^2 b^3 = E\mu^2 e^{i\omega t} \quad (39)$$

within $O(\mu^4)$ with

$$G(D) \equiv f(D) + c_1 g(D).$$

The first term may be written

$$\begin{aligned} e^{i\omega t} G(D+i\omega) b e^{i\phi} &= e^{i\omega t} \{G(i\omega) + G'(i\omega)D\} b e^{i\phi} \\ &= G(i\omega) b e^{i(\omega t + \phi)} + G'(i\omega) (\dot{b} + i b \dot{\phi}) e^{i(\omega t + \phi)} \\ &= |\epsilon| A e^{i\alpha} (\delta - i\sigma) b e^{i(\omega t + \phi)} - A e^{i\alpha} (\dot{b} + i b \dot{\phi}) e^{i(\omega t + \phi)}. \end{aligned}$$

Substitute this expression into (39), putting $|\epsilon| = \mu^2$, separate real and imaginary parts after dividing by $\mu^2 A e^{i\alpha}$, and the variational equations (37), (38) are obtained immediately.

12. General remarks on the method

(i) The approximate linear equation has been taken as

$$\{(D - \epsilon)^2 + \Omega^2\} h(D) x = 0. \quad (40)$$

Neither this form, nor the normalization to $\Omega = 1$ is essential. If the approximate linear equation is

$$\{(D^2 + \Omega_1^2) h(D) + \epsilon_1 k(D)\} x = 0 \quad (41)$$

the same procedure may be used, as is done in (2) with Ω_1 normalized to 1. If we take $z = i\Omega_1$ as an approximate root of

$$(z^2 + \Omega_1^2) h(z) + \epsilon_1 k(z) = 0$$

we obtain for the true value of the root

$$z = i\Omega_1 - \frac{\epsilon_1 k(i\Omega_1)}{2i\Omega_1 h(i\Omega_1)} + O(\epsilon_1^2).$$

If this is written

$$z = i\Omega_1 + \epsilon_1 (M + iM')$$

within $O(\epsilon_1^2)$ we shall have

$$\epsilon = \epsilon_1 M$$

and

$$\Omega = \Omega_1 + \epsilon_1 M'.$$

With

$$\epsilon_1 \sigma_1 = \frac{\omega - \Omega_1}{\Omega_1} \quad \text{and} \quad \epsilon \sigma = \frac{\omega - \Omega}{\Omega}$$

we obtain

$$\epsilon \sigma = \epsilon_1 \left(\sigma_1 - \frac{M'}{\Omega} \right).$$

Thus with (41) instead of (40) the ratio $\epsilon/\epsilon_1 = M$ enters the equations as a scale factor, and a change of origin of σ_1 to $-M'/\Omega$ is required to bring the variational equations to the standard form. These equations are also obtained by the symbolic procedure of the previous section.

(ii) The relative importance of the non-linear terms has been related, in the natural way, to the position of the term in the Taylor expansion of y in powers of x . A different assignment of the relative importance of these terms could have been made, and the expansions carried out in powers of μ instead of in powers of B or b . The procedure would have been essentially the same, though the results would have lost some of their symmetry.

(iii) The case $\epsilon = 0$ has not been mentioned explicitly. The non-resonant solution fails completely in this case but the resonant solution may be used.

(iv) Fundamental resonance has been considered, but the method may be adapted to other cases; e.g. if the forcing term were $2B \cos 2\omega t$, with $\omega - 1 = \epsilon\sigma$, and $x^{(1)} = 2b \cos(\omega t + \phi)$, equations for subharmonic resonance of order 2 could have been derived.

(v) An upper limit to the error made by terminating the series at any particular term may be obtained from the remainder after the same number of terms of the majorante series.

(vi) In the case when the degree of $g(z)$ is higher than that of $f(z) + c_1 g(z)$, the problem becomes a singular perturbation problem. The solution fails because $\zeta(z)$ tends to zero at infinity and it is no longer possible to choose N less than $|\zeta(ri\omega)|$ for all r . If, however, $|\zeta(ri\omega)|$ remains greater than C_1 for $r \leq r_1$, say, then with $N = C_1$ we may develop the solution as far as the terms containing μ^{2n} where $2n \leq r_1$. If b and ϕ satisfy the amplitude equation as far as this term, a solution will be obtained which satisfies the differential equation with a residual less than N times the remainder after this number of terms of the majorante series, which may be very small if r_1 is large.

REFERENCES

1. A. W. GILLIES, 'The periodic solutions of a non-linear differential equation of the second order with unsymmetrical non-linear damping and a forcing term', *Quart. J. Mech. App. Math.* **8** (1955) 108.
2. — 'The periodic solutions of the differential equation of a resistance-capacitance oscillator', *ibid.* **10** (1957) 101.

RESPONSE FUNCTIONS OF LINEAR SYSTEMS WITH CONSTANT COEFFICIENTS HAVING ONE DEGREE OF FREEDOM†

By A. W. BABISTER (*James Watt Eng. Lab., The University, Glasgow*)

[Received 4 June 1956]

SUMMARY

The transient response is considered for systems satisfying linear differential equations with constant coefficients. Criteria for optimizing the response are given in terms of the response functions

$$L = \int_0^{\infty} e^2 d\tau \quad \text{and} \quad L_1 = \int_0^{\infty} \left(\frac{de}{d\tau} \right)^2 d\tau,$$

where e is the error at time τ . Expressions are obtained for both L and L_1 in terms of (a) the coefficients of the characteristic equation, and (b) the frequency response of the system. It is shown how the response due to (i) a step function disturbance, (ii) an initial impulse, and (iii) a constant velocity input can be simply related to the response in the free motion.

1. Response functions and optimum response

WHEN considering the performance of a system (e.g. a dynamical system or servomechanism) we are interested in the accuracy with which the output of the system follows the input. More precise elaboration of this general statement depends upon the particular application, the order of accuracy and the sensitivity of the system varying greatly with different applications. The systems considered may be mechanical, electrical, hydraulic, or aerodynamic. The terms used here (e.g. forces, equations of motion) are mainly based on mechanical systems, but the analysis is, of course, perfectly general.

We shall be concerned, in general, with stable systems, i.e. systems which when subjected to any disturbance acting for a finite time ultimately return to their initial state. The performance of such a system is intuitively measured by such factors as its overshoot when subjected to a step disturbance, the oscillatory nature and damping in that case (the transient motion), and also by the amplitude of the motion and the resonance peak and frequency when the system is subjected to a steady sinusoidal disturbing force (the frequency response of the motion). Often the designer has freedom to choose the value of many of the parameters of the system

† This paper forms part of a thesis for which the Ph.D. degree of the University of Glasgow was awarded.

(e.g. degree of damping, spring stiffness). The problem of optimization is the selection of such values of these variables that the response of the system is 'the best' or, more often, the most satisfactory for the particular application. The optimum values will depend on the particular input disturbance. Thus at the start we are confronted with the choice of basing our analysis either on the transient behaviour of the system or on its frequency response. The nature of this fundamental choice will depend on which is the most likely type of input the system will have to deal with. Fortunately many systems having a satisfactory transient also have a satisfactory frequency response. This is not surprising since the response of a system to any disturbance is governed by the differential equation of motion of the system. In fact the transient and frequency response can be correlated by a Fourier transformation. Empirical relations are often used, based, for example, on the relationship between the peak overshoot and the resonance peak, or between the resonant frequency and the transient oscillatory peak, or between the resonant frequency and the speed of response (see (1)).

We shall endeavour to give simple mathematical criteria for optimizing the transient response. Many such criteria have been used in recent years, mainly for systems subjected to input step disturbances. In references (2) and (3) the integrals

$$I_1 = \int_0^{\infty} e \, d\tau \quad \text{and} \quad I_2 = \int_0^{\infty} \tau e \, d\tau$$

(where e is the error) are minimized. These criteria break down when the response is oscillatory, since an overshoot decreases the value of I_1 and I_2 . In (4) and (5) the criterion for optimizing the transient response is that

$$I_3 = \int_0^{\infty} e^2 \, d\tau$$

should be a minimum. This criterion is widely used and can be readily handled, either analytically or by a computing machine. However, in general it leads to a slightly overdamped response often with a large undesirable overshoot. To overcome this Mack considers (6) the integral

$$I_4 = \int_0^{\infty} \tau^2 e^2 \, d\tau.$$

The system making I_4 a minimum gives a very satisfactory performance, but the formula for I_4 is often troublesome to evaluate.

In reference (7)

$$I_5 = \int_0^{\infty} \tau |e| \, d\tau,$$

the integral of time-multiplied absolute-value of error (ITAE) is minimized. The ITAE criterion works very well for response of zero-displacement-error systems to a step input, but not so well for zero-velocity-error systems. It may indeed be thought to be over selective in its choice of optimum, distinguishing too sharply between the optimum and systems near the optimum. From the form of the integral it allows sizeable errors for small values of τ ; this is seen especially in its choice of optimum response for zero-velocity-error and zero-acceleration-error systems. While I_5 is easy to calculate, it is difficult to handle analytically, and thus it is hard (if not impossible) to extend results of the ITAE criterion to high order differential equations.

We shall consider here the response functions

$$L = \int_0^{\infty} e^2 d\tau \quad \text{and} \quad L_1 = \int_0^{\infty} \left[\frac{de}{d\tau} \right]^2 d\tau.$$

As stated above a system based on minimizing L often has an undesirable overshoot. Often L has a rather flat minimum as the parameters of the system are varied, the value of this minimum being relatively insensitive to small changes in some of the parameters. The system having the most satisfactory performance is impossible to define precisely in mathematical terms. For a system with a large overshoot $de/d\tau$ will be large at some time and thus L_1 will be correspondingly large. Our aim will therefore be to choose values of the parameters so that L is near its minimum value while L_1 is considerably lower than its value at L_{\min} . These systems will be of a less oscillatory nature than those given by the mean square optimum.

2. Derivation of simple formulae for the response functions L and L_1

We consider a system for which the equation of motion is

$$a_n D^n x + a_{n-1} D^{n-1} x + a_{n-2} D^{n-2} x + \dots + a_1 D x + a_0 x = f(\tau), \quad (1)$$

where

$$D \equiv \frac{d}{d\tau},$$

$a_0, a_1, a_2, \dots, a_n$ are constants ($a_n \neq 0$) and $f(\tau)$ is some known function. The initial conditions are given by

$$x = x_0, \quad Dx = Dx_0, \quad D^2x = D^2x_0, \quad \text{etc.}, \quad D^{n-1}x = D^{n-1}x_0 \quad \text{at } \tau = 0.$$

The characteristic equation corresponding to (1) is

$$F(\lambda) = a_n \lambda^n + a_{n-1} \lambda^{n-1} + a_{n-2} \lambda^{n-2} + \dots + a_0 = 0. \quad (2)$$

We shall be concerned only with stable systems for which all the roots of (2) are negative or have negative real parts.

If there were no lag in the system x would equal $(1/a_0)f(\tau)$ at all times. The error e is given by

$$e = \text{input} - \text{output} = \frac{1}{a_0}f(\tau) - x. \quad (3)$$

As stated above we shall derive formulae for

$$L = \int_0^{\infty} e^2 d\tau = \int_0^{\infty} \left[x - \frac{1}{a_0}f(\tau) \right]^2 d\tau \quad (4)$$

$$L_1 = \int_0^{\infty} \left[\frac{de}{d\tau} \right]^2 d\tau = \int_0^{\infty} \left[\frac{dx}{d\tau} - \frac{1}{a_0}f'(\tau) \right]^2 d\tau. \quad (5)$$

We shall now consider various forms of the functions $f(\tau)$ and the corresponding formulae for the response functions L and L_1 .

3. Free motion: $f(\tau) = 0$

We write
$$L = \int_0^{\infty} x^2 d\tau, \quad L_s = \int_0^{\infty} [D^s x]^2 d\tau. \quad (6)$$

Putting $f(\tau) = 0$ and multiplying (1) successively by $x, Dx, \dots, D^{n-1}x$ and integrating from 0 to ∞ we obtain n simultaneous linear equations for L, L_1, \dots, L_{n-1} in terms of $x_0, Dx_0, \dots, D^{n-1}x_0$. We find

$$L = \frac{1}{a_0 T_{n-1}} \begin{vmatrix} \beta_1 & a_2 & a_4 & a_6 & . & . & . \\ -\beta_2 & a_1 & a_3 & a_5 & . & . & . \\ \beta_3 & a_0 & a_2 & a_4 & . & . & . \\ -\beta_4 & 0 & a_1 & a_3 & . & . & . \\ . & . & . & . & . & . & . \end{vmatrix} \quad (7)$$

$$L_1 = \frac{1}{T_{n-1}} \begin{vmatrix} \beta_2 & a_3 & a_5 & a_7 & . & . & . \\ -\beta_3 & a_2 & a_4 & a_6 & . & . & . \\ \beta_4 & a_1 & a_3 & a_5 & . & . & . \\ -\beta_5 & a_0 & a_2 & a_4 & . & . & . \\ . & . & . & . & . & . & . \end{vmatrix} \quad (8)$$

where T_{n-1} is the $(n-1)$ th test function (see (8) and (9)) given by

$$T_{n-1} = \begin{vmatrix} a_1 & a_3 & a_5 & a_7 & . & . & . \\ a_0 & a_2 & a_4 & a_6 & . & . & . \\ 0 & a_1 & a_3 & a_5 & . & . & . \\ 0 & a_0 & a_2 & a_4 & . & . & . \\ . & . & . & . & . & . & . \end{vmatrix} \quad (9)$$

and the determinants in (7) and (8) are of the n th order while the determinant in (9) is of order $(n-1)$.

The β 's are quadratic functions in the initial conditions $x_0, \dots, D^{n-1}x_0$. Thus

$$\begin{aligned}\beta_1 = & x_0(\tfrac{1}{2}a_1x_0 + a_2Dx_0 + a_3D^2x_0 + a_4D^3x_0 + \dots) - \\ & - Dx_0(\tfrac{1}{2}a_3Dx_0 + a_4D^2x_0 + a_5D^3x_0 + a_6D^4x_0 + \dots) + \\ & + D^2x_0(\tfrac{1}{2}a_5D^2x_0 + a_6D^3x_0 + a_7D^4x_0 + a_8D^5x_0 + \dots) - \\ & - D^3x_0(\tfrac{1}{2}a_7D^3x_0 + a_8D^4x_0 + a_9D^5x_0 + a_{10}D^6x_0 + \dots) + \dots, \quad (10)\end{aligned}$$

$$\begin{aligned}\beta_2 = & \tfrac{1}{2}a_0x_0^2 + Dx_0(\tfrac{1}{2}a_2Dx_0 + a_3D^2x_0 + a_4D^3x_0 + \dots) - \\ & - D^2x_0(\tfrac{1}{2}a_4D^2x_0 + a_5D^3x_0 + a_6D^4x_0 + \dots) + \\ & + D^3x_0(\tfrac{1}{2}a_6D^3x_0 + a_7D^4x_0 + a_8D^5x_0 + \dots) - \dots, \quad (11)\end{aligned}$$

$$\begin{aligned}\beta_3 = & Dx_0(a_0x_0 + \tfrac{1}{2}a_1Dx_0) + D^2x_0(\tfrac{1}{2}a_3D^2x_0 + a_4D^3x_0 + a_5D^4x_0 + \dots) - \\ & - D^3x_0(\tfrac{1}{2}a_5D^3x_0 + a_6D^4x_0 + a_7D^5x_0 + \dots) + \\ & + D^4x_0(\tfrac{1}{2}a_7D^4x_0 + a_8D^5x_0 + a_9D^6x_0 + \dots) - \dots, \quad (12)\end{aligned}$$

$$\begin{aligned}\beta_4 = & -\tfrac{1}{2}a_0Dx_0^2 + D^2x_0(a_0x_0 + a_1Dx_0 + \tfrac{1}{2}a_2D^2x_0) + \\ & + D^3x_0(\tfrac{1}{2}a_4D^3x_0 + a_5D^4x_0 + a_6D^5x_0 + \dots) - \\ & - D^4x_0(\tfrac{1}{2}a_6D^4x_0 + a_7D^5x_0 + a_8D^6x_0 + \dots) + \\ & + D^5x_0(\tfrac{1}{2}a_8D^5x_0 + a_9D^6x_0 + a_{10}D^7x_0 + \dots) - \dots. \quad (13)\end{aligned}$$

These equations simplify considerably when the initial values of all except one $D^m x$ are zero.

As shown in (8) and (9), for stability (with $a_n > 0$) all the test functions must be positive. The critical stability criteria are $T_{n-1} > 0$ and $a_0 > 0$. When a_0 vanishes, the characteristic equation (2) has a zero root and the system is neutrally stable. When T_{n-1} vanishes, the system has a pair of equal and opposite roots. In both of these cases if the system is displaced from its original position it would never return to and remain in that position; thus L is infinite if either a_0 or T_{n-1} vanish, as shown by (7).

We see that L and L_1 are both second degree expressions in the initial conditions. An alternative expression for L_1 can be obtained directly from L by replacing x_0 by Dx_0 , Dx_0 by $D^2x_0, \dots, D^{n-1}x_0$ by $D^n x_0$ in the expression for L .

4. Derivation of integral formulae for the response functions in terms of the frequency response spectrum of the system

As stated above, the transient and the frequency response can be correlated by a Fourier transformation. It is of interest to relate the response functions to the amplitude and phase of the frequency response characteristics of the system.

As shown in (10), if

$$\left. \begin{aligned} G(\omega) &= \int_{-\infty}^{\infty} e^{-i\omega\tau} x(\tau) d\tau, \\ x(\tau) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega\tau} G(\omega) d\omega, \end{aligned} \right\} \quad (14)$$

provided that $\int_{-\infty}^{\infty} x(\tau) d\tau$ is absolutely convergent. We shall consider only the special case in which $x(\tau) = 0$ for $\tau < 0$.

In the free motion we have

$$x = \sum_{r=1}^n A_r e^{\lambda_r \tau}, \quad (15)$$

where A_r ($r = 1$ to n) are determined by the initial conditions and λ_r is a root of the characteristic equation (2). From (14) and (15) we have

$$G(\omega) = \sum \frac{A_r}{i\omega - \lambda_r},$$

provided the system is stable. On substituting for A_r we find

$$G(\omega) = \frac{g_0 + g_1(i\omega) + g_2(i\omega)^2 + \dots + g_{n-1}(i\omega)^{n-1}}{a_0 + a_1(i\omega) + a_2(i\omega)^2 + \dots + a_n(i\omega)^n}, \quad (16)$$

where

$$g_s = a_{s+1}x_0 + a_{s+2}Dx_0 + \dots + a_n D^{n-s-1}x_0 \quad (s = 0 \text{ to } n-1). \quad (17)$$

From Parseval's theorem, for stable systems,

$$\frac{1}{\pi} \int_0^{\infty} |G(\omega)|^2 d\omega = \int_0^{\infty} x^2(\tau) d\tau,$$

since $x(\tau)$ is real.

Thus, in the free motion

$$L = \frac{1}{\pi} \int_0^{\infty} \frac{[g_0 - g_2\omega^2 + g_4\omega^4 - \dots]^2 + \omega^2[g_1 - g_3\omega^2 + \dots]^2}{[a_0 - a_2\omega^2 + a_4\omega^4 - \dots]^2 + \omega^2[a_1 - a_3\omega^2 + \dots]^2} d\omega, \quad (18)$$

where g_s is given by (17), $s = 0$ to $n-1$.

Equation (18) is an extension of the formula given in (6). For sufficiently large values of ω the integrand tends to $g_{n-1}/(a_n \omega)^2$. Thus the integral converges like $1/\omega$.

Now from (14) we see that $|G(\omega)|$ represents the frequency spectrum of the given system. If the system is given a periodic disturbance $a_0 e^{i\omega\tau}$, we find from (1) that

$$x = \frac{a_0 e^{i\omega\tau}}{a_0 + a_1 i\omega - a_2 \omega^2 - a_3 i\omega^3 + \dots}.$$

The amplitude (or dynamic magnification) M and phase advance N are given by

$$\frac{1}{M^2} = \frac{1}{a_0^2} \{ [a_0 - a_2 \omega^2 + a_4 \omega^4 - \dots]^2 + \omega^2 [a_1 - a_3 \omega^2 + \dots]^2 \} \quad (19)$$

and

$$N = -\tan^{-1} \left(\frac{a_1 \omega - a_3 \omega^3 + a_5 \omega^5 - \dots}{a_0 - a_2 \omega^2 + a_4 \omega^4 - \dots} \right). \quad (20)$$

We see that the expressions on the right-hand sides of equations (19) and (20) are precisely those that occur in the denominator of (18). Large amplitudes will occur when the damping is small and the forcing frequency is close to one of the natural frequencies of the system. From (19) it is seen that if the dynamic magnification is large over a range of frequencies, the integrand in (18) would be expected to be correspondingly large and L would be well away from its optimum value.

5. Response of a system with no initial displacement

We consider a system with initial conditions

$$x_0 = Dx_0 = D^2x_0 = \dots = D^{n-1}x_0 = 0, \quad (21)$$

the system satisfying the equation of motion (1).

(i) *Step function disturbance*

$$f = 0 \quad \text{for } \tau \leq 0; \quad f = f_0 \quad \text{for } \tau > 0,$$

where f_0 is a constant. Writing

$$x' = x - \frac{f_0}{a_0}$$

we see from (1) that the response of the system following a step disturbance is identical in form with that in the free motion for which

$$x_0 = -\frac{f_0}{a_0} \quad \text{and} \quad Dx_0, D^2x_0, \dots, D^{n-1}x_0 \quad \text{are zero.}$$

The errors in the two responses are the same at any given time and the values of L and L_1 can be found from (7) and (8). Thus

$$L = \frac{f_0^2}{2a_0^3 T_{n-1}^2} \begin{vmatrix} a_1 & a_2 & a_4 & a_6 & \cdot & \cdot & \cdot \\ -a_0 & a_1 & a_3 & a_5 & \cdot & \cdot & \cdot \\ 0 & a_0 & a_2 & a_4 & \cdot & \cdot & \cdot \\ 0 & 0 & a_1 & a_3 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{vmatrix}, \quad (22)$$

$$L_1 = \frac{f_0^2}{2a_0 T_{n-1}^2} \begin{vmatrix} a_2 & a_4 & a_6 & a_8 & \cdot & \cdot & \cdot \\ a_1 & a_3 & a_5 & a_7 & \cdot & \cdot & \cdot \\ a_0 & a_2 & a_4 & a_6 & \cdot & \cdot & \cdot \\ 0 & a_1 & a_3 & a_5 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{vmatrix}. \quad (23)$$

It follows that the system having the most satisfactory response to a step function will have the most satisfactory response in the corresponding free motion.

(ii) *Response to an initial unit impulse*

It is readily shown that just as the unit step function is the integral of the unit impulse, so the response to a unit step disturbance is the integral of the response to a unit impulse. The response functions can thus be obtained directly from the response functions for a step function disturbance. We find

$$L_s \text{ (unit impulse)} = L_{s+1} \text{ (unit step disturbance)}, \quad (24)$$

where $L_0 \equiv L$.

(iii) *Constant velocity input*

$$f = 0 \quad \text{for } \tau \leq 0; \quad f = f_1 + f_0 \tau \quad \text{for } \tau > 0,$$

where f_0 and f_1 are constants.

When the transient has died away, the steady motion is given by

$$x = a\tau + b, \quad (25)$$

$$\text{where, from (1),} \quad f_0 = a_0 a \quad \text{and} \quad f_1 = a_1 a + a_0 b. \quad (26)$$

Writing

$$x' = x - a\tau - b$$

we see from (1) that the response of the given system is identical in form with that in a free motion with $x_0 = -b$, $Dx_0 = -a$, and $D^2x_0, \dots, D^{n-1}x_0$ zero. In particular when $f_0/a_0 = f_1/a_1$, $b = 0$ and the equivalent free system has $Dx_0 = -f_0/a_0$ and $x_0, D^2x_0, \dots, D^{n-1}x_0$ zero. We see that such a system ultimately has no static error since $x' \rightarrow 0$ and there is no position error with constant velocity input (since $b = 0$). From (7),

$$L = \frac{f_0^2}{2a_0^3 T_{n-1}} \begin{vmatrix} -a_3 & a_2 & a_4 & a_6 & \cdot & \cdot & \cdot \\ -a_2 & a_1 & a_3 & a_5 & \cdot & \cdot & \cdot \\ a_1 & a_0 & a_2 & a_4 & \cdot & \cdot & \cdot \\ a_0 & 0 & a_1 & a_3 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{vmatrix}. \quad (27)$$

Similar formulae can be deduced for a constant acceleration input.

6. Extensions to the above analysis

As is well known, the response of a linear system to an arbitrary disturbance can be simply related to its response to unit impulses. When the input is expressed as a Fourier integral the response functions L and L_1 can be found by an extension of the method given above, using Parseval's theorem. The response functions for a finite square wave disturbance could be obtained in this manner.

7. Response functions for unstable systems

The response functions defined above apply only to stable systems. It sometimes happens that the system possesses one slight unstable mode (e.g. an aircraft with spiral instability). In this case we are primarily interested in the behaviour of the system shortly after the disturbance takes place. The response functions can be used to give a measure of the response of the system, the upper limits of the integrals for L and L_1 being taken to be some convenient finite time.

REFERENCES

1. G. J. THALER and R. G. BROWN, *Servomechanism Analysis* (McGraw-Hill, 1953).
2. R. C. OLDENBOURG and H. SARTORIUS, *The Dynamics of Automatic Controls* (ASME, 1948).
3. P. T. NIMS, 'Some design criteria for automatic controls', *Trans. AIEE* **70** (1951), 606.
4. N. WIENER, *Extrapolation, Interpolation and Smoothing of Stationary Time Series* (Wiley, 1950).
5. A. C. HALL, *The Analysis and Synthesis of Linear Servomechanisms* (MIT, 1943).
6. C. MACK, 'The calculation of the optimum parameters for a following system', *Phil. Mag.* (7) **40** (1949), 922.
7. D. GRAHAM and R. C. LATHROP, 'The synthesis of optimum transient response: criteria and standard forms', *Trans. AIEE* **72** (1953), 273.
8. A. HURWITZ, 'Ueber die Bedingungen, unter welchen eine Gleichung nur Wurzeln mit negativen reellen Theilen besitzt', *Math. Ann.* **46** (1895), 273.
9. R. A. FRAZER and W. J. DUNCAN, 'On the criteria for the stability of small motions', *Proc. Roy. Soc. London*, A **124** (1929), 642.
10. H. S. CARSLAW and J. C. JAEGER, *Operational Methods in Applied Mathematics* (Oxford, 1941).

AN ELECTROLYTIC TANK AS AN ANALOGUE COMPUTING MACHINE FOR FACTORIZING HIGH DEGREE POLYNOMIALS

By S. K. IP (*Imperial College, London*)

[Received 25 June 1954; revise received 15 January 1956]

SUMMARY

The analogue representation of a polynomial by means of a potential distribution set up by point electrodes in a uniform sheet of conducting material was first described by Lucas. The electrodes are suitably placed along the real axis of the complex z -plane after a region in which the roots may be expected to lie has been determined. The roots can then be found by exploring the electrolytic tank surface for points of zero potential gradient. In addition to the double-layer type which is described fully by Cherry *et al.*, a single-layer transformed electrolytic tank, representing the infinite z -plane, was employed throughout the problems explored as a basis for this investigation. By means of simple measurements, all the roots of a polynomial as high as the 16th degree could be located in between 2 to 4 hours' time with an overall accuracy of about 4 per cent. For special cases a method is given for setting up the potential field by electrodes placed along the imaginary axis.

1. The physical representation of algebraic functions

1.1. The algebraic function analogy as described by Lucas

Let us consider the equation

$$f(z) \equiv a_n z^n + a_{n-1} z^{n-1} + \dots + a_1 z + a_0 = 0 \quad (1)$$

of degree n with real coefficients and let us take $n+1$ arbitrary unequal quantities x_1, x_2, \dots, x_{n+1} , each determining a point L on the real x -axis of the complex z -plane. Let us now form the auxiliary polynomial

$$F(z) = \prod_{s=1}^{n+1} (z - x_s) \quad (2)$$

where $F(z)$ is of higher order than $f(z)$.

By dividing $f(z)$ by $F(z)$, where the coefficients of each are real, we obtain a rational function $G(z)$ which can be expanded in partial fractions:

$$G(z) = \sum_{s=1}^{n+1} \frac{k_s}{z - x_s} \quad (3)$$

The residues k_1, k_2 , etc., can easily be determined. They are all real, some positive and some negative.

To determine the positions of the roots of $f(z)$, the complex plane of $z = x + iy$ is represented on a uniformly conducting sheet (electrolyte),

and each of the partial fractions is represented by feeding a current of strength k_s into the sheet at the positions x_s . A system of current flow is then set up which characterizes the function $G(z)$.

1.2. Conditions at a root point in a potential field

Let us consider a point $P(z)$ in a potential distribution

$$V(z) = \sum_{s=1}^{n+1} k_s \log(z - \xi_s) \quad (4)$$

due to a number of line sources of strength k_s at positions ξ_s . The condition that $V(z)$ be 'stationary' near a point z_0 is $dV/dz_0 = 0$, that is,

$$\sum \frac{k_s}{(z_0 - \xi_s)} = 0$$

which can be seen from equation (3) to be identical with $G(z) = 0$. This same condition constitutes the solution of $f(z) = 0$ in (1).

Then if $G(z)$ be thus represented by means of the gradient of a potential distribution, the positions of the points z_0 will be the roots of the proposed equation $f(z) = 0$.

The positions of the points z_0 in the potential field can be traced out by means of a single probe. This makes use of the fact that at the point z_0 where $dV/dz = 0$, i.e. the point with zero potential gradient, there is a saddle point of potential, as follows from Dirichlet's theorem.

To investigate the behaviour of the equipotentials in the vicinity of z_0 let us put

$$z = (z_0 + \Delta z)$$

so that

$$V(z) = V(z_0 + \Delta z). \quad (5)$$

Then expanding (5) in a Taylor series, we have

$$V(z_0 + \Delta z) = V(z_0) + \Delta z V'(z_0) + \frac{1}{2}(\Delta z)^2 V''(z_0) + O(\Delta z)^3. \quad (6)$$

Hence the equipotentials in the vicinity of z_0 are given approximately by

$$(\Delta z)^2 = \text{constant},$$

that is, either the hyperbolas

$$(\Delta x)^2 - (\Delta y)^2 = \text{constant}$$

or

$$\Delta x \Delta y = \text{constant}.$$

Thus equipotential lines around z_0 occur in hyperbolic form. The limiting equipotentials (the asymptotes of these sets of hyperbolas) are straight lines that intersect at z_0 . Moreover, since z_0 is a saddle point, AB and CD in Fig. 1 are loci of potential minima and maxima respectively, intersecting at the root point.

On the electrolytic tank, current-carrying electrodes of small diameter set up a distribution of potential which is peculiar to the given polynomial.

The potential maxima and minima are traced out; their intersections will be the positions of the root points of the polynomial.

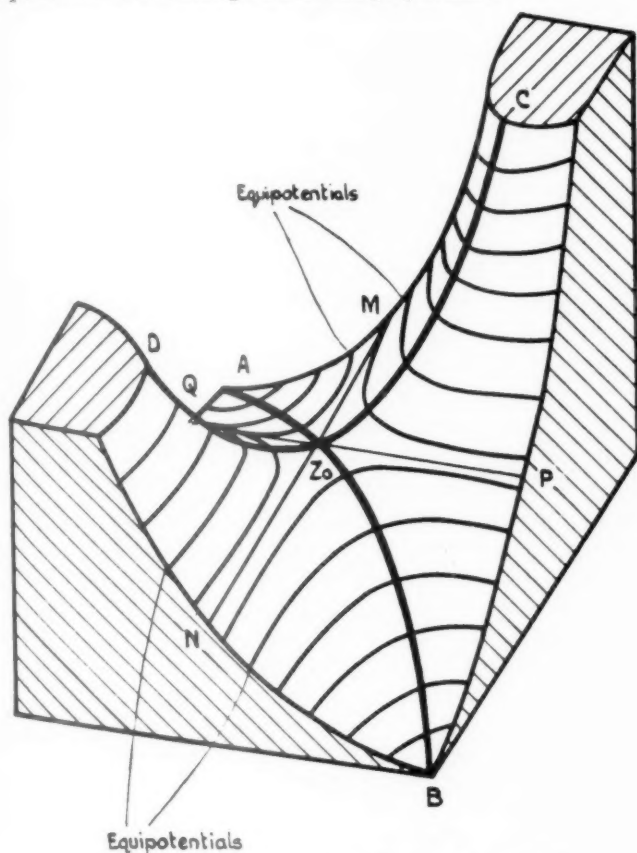


FIG. 1. Sectional view of a saddle point of potential

2. The electrolytic tank method

It is often extremely difficult to factorize polynomials exactly, but there are approximate methods for numerical solutions (2)–(8).

The following is an electrical method based on the practical experience of the electrolytic tank analogue found by Boothroyd, Cherry, and Makar (9), and is a technique of the greatest value for representing a complex plane. It is applicable, in many cases, for solving both the real and complex roots of rational algebraic equations of any order that have real coefficients. The equipment employed is simple and inexpensive compared with other

mechanical and electrical computing machines (10), (11), (12), and furthermore is easy to operate.

2.1. The single-layer electrolytic tank

By the bilinear transformation, the whole z -plane above the real axis is conformally transformed into a unit circle, and is represented by the region within the single-layer tank. The relation between z and the new complex variable ζ is defined by

$$\zeta = -\frac{1}{z+i}. \quad (7)$$

In practical form, a Perspex ring 1 cm thick and 40 cm inner diameter, fixed on to a plate-glass disk, forms the wall of the single-layer electrolytic tank. The real axis ($-\infty$ to ∞) of the z -plane now becomes a circle of radius $\frac{1}{2}$ and centre at $\frac{1}{2}i$ in the ζ -plane, and forms the boundary of the tank. Since the complex roots of a polynomial with real coefficients occur in conjugate pairs we need deal only with one-half of the z -plane (either above or below the real axis) and the tank gives us the whole of such a region. Note that the coordinates which are marked on the plate-glass must be curvilinear instead of rectangular.

Copper sulphate solution (concentration 4 to 5 g to 1 l. of distilled water) to a depth of 1 cm is used as electrolyte. Operating currents, of the order of milliamps, are fed from a 500 c/s stabilized current source; positive and negative currents can be taken as being in opposite phase to each other.

In cases where the roots have large numerical values, or all are offset to one side of the imaginary axis, we can bring them into the favourable region 2 to -2 on the real axis by a suitable linear transformation. There is no need to apply this transformation to the equation itself, but only to the roots when they are found.

For example, suppose the roots lie between $x = -150$ to -50 . Then the transformation is

$$z' = \frac{z+100}{25},$$

so that a root z_r in the coordinates of the tank gives $25z'_r - 100$ as a root of the equation.

If we put the electrodes at $-150, -125, -100, -75, -50$, this would imply that they are placed at the points $-2, -1, 0, 1, 2$ in the z' case.

2.2. The positioning of the electrodes

The positioning of the electrodes x_1, x_2 , etc., is arbitrary. But, since the electrode field intensity at a point is inversely proportional to the distance between the point and the electrodes, the latter should be placed as near to the root points as possible. If it is possible to find a limiting region S on the

real axis, inside which lie all possible roots of a polynomial, this will be the optimum region in which to place the electrodes.

One method of determining the region S is by plotting $f(z)$ against the real x -axis: in the vicinity of the roots, the variation of the function is more gradual than it is at a distance from the roots. When a polynomial is made free from real roots, as was generally done (by standard methods) so as to reduce the polynomial to the lowest possible degree, the graph will not touch the real axis and the region S can be taken as lying between the two steep slopes of the graph as shown in Fig. 2 (a).

After the region S is determined, $n+1$ real values x_1, x_2 , etc., are arbitrarily chosen from it (section 1.1). When possible, it is desirable that these $n+1$ values should be distributed evenly within the region. Further, they should be convenient values, otherwise unnecessary complications and labour may be entailed in the solution of the partial fractions that follow.

When the roots of a polynomial occur alongside the imaginary axis, $f(z)$ can be formed into a new function $H(z) = f(z)f(-z)$. By plotting $H(z)$ against the imaginary axis, a similar limiting region S can be obtained on the imaginary axis. In this way potential distribution can be set up by electrodes along the imaginary axis.

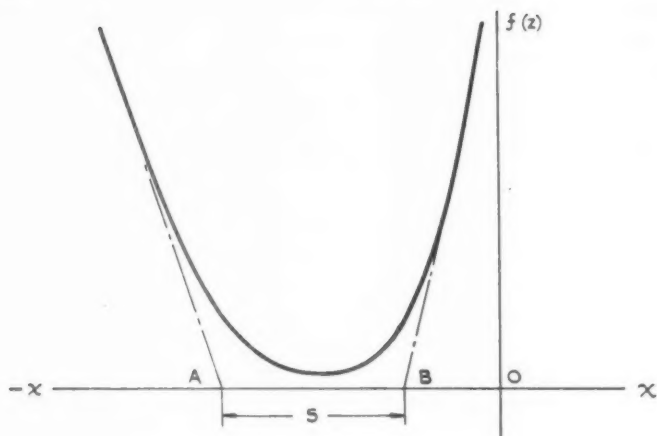
A general method for obtaining the boundary values of the roots is to plot $f(z)$ against both the real and imaginary axes; hence the roots can be located approximately within the rectangle bounded by the four limiting points A, B, C , and D , as shown in Fig. 2 (b).

2.3. *A quick and accurate method of searching for the root points of a polynomial on the electrolytic tank*

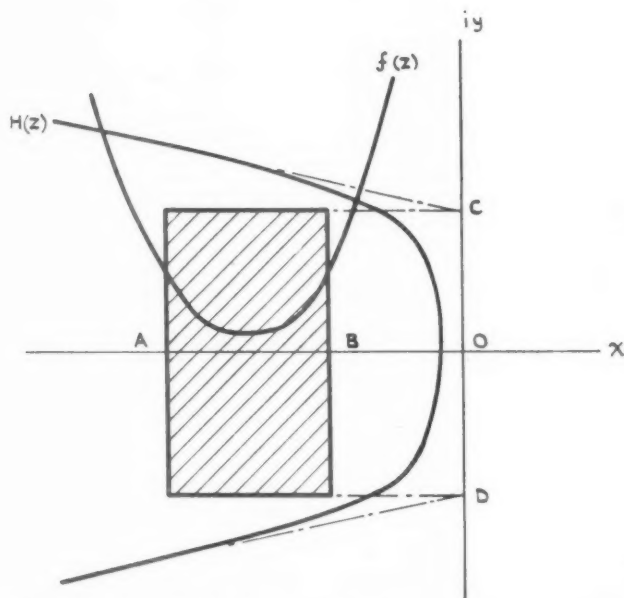
On the electrolytic tank a single moving probe is used to measure potentials with respect to an arbitrary point on the tank surface. The probe sweeps the tank, firstly in closely-spaced parallel lines and then similarly in the normal direction, until the whole area of the tank has been explored.

The potential at various positions can meanwhile be observed on the galvanometer. When the potential reaches either a maximum or a minimum, the position on the tank surface is recorded on a sheet of graph paper usually having the same size as the tank surface. At the end of the search throughout the tank surface, the locations of the saddle points can immediately be found where two loci of potential maxima and minima intersect.

Then within appropriate regions in the vicinity of the intersecting points equipotential lines are carefully traced out. The saddle points of the equi-



(a)



(b)

FIG. 2. Approximating the locations of the complex roots of a polynomial by an area bounded by four limiting points

potential lines in cartesian coordinates will give the complex values of the roots of the polynomial represented by that potential distribution. In practice, the position of a saddle-point cannot be located precisely by means of potential field configuration; however, a null region is found and the exact position of the saddle point can thus be taken at the intersection of the asymptotes of the equipotentials as shown in Fig. 1.

- (a) Care must be taken when searching for the potential saddle points. For example, Fig. 1 illustrates a sectional view of a saddle point of potential. If in this case the probe travels along Mz_0N and Pz_0Q , that is, along an equipotential, neither a maximum nor a minimum will be recorded. It is sometimes necessary to make further explorations in a direction of 45° from each of the original paths of probe movement, that is, along Az_0B and Cz_0D .

3. Practical examples

The two examples given below were solved by the transformed single-layer electrolytic tank, on which electrodes were positioned along the real (Example 1) and imaginary (Example 2) axes.

Example 1 is due to Hitchcock (7) and was also solved by Lin (8). The roots are well distributed, which is one of the most favourable conditions for a solution by means of the electrolytic tank. A close degree of accuracy was obtained.

By placing the current-carrying electrodes along the imaginary axis, Example 2 was solved with the same procedure given in Section 2, in spite of the high order of the function.

Example 1

Here

$$f(z) = z^8 - 3.012z^7 + 3.225z^6 + 1.0212z^5 + 6.9862z^4 - 21.889z^3 + 8.110z^2 + 5.9012z + 23.889; \quad (8)$$

$f(z)$ is plotted against the real axis in Fig. 3 from which the electrode region S is chosen between -2.0 and 2.0 with the x 's at $2.0, 1.5, 1.0, 0.5, 0, -0.5, -1.0, -1.5, -2.0$. The values of the residues are found to be

$k_1 = 0.726012$	$k_4 = -9.468622$	$k_7 = 10.877688$
$k_2 = -1.591714$	$k_5 = 10.613733$	$k_8 = -12.626996$
$k_3 = 4.485866$	$k_6 = -9.310044$	$k_9 = 7.280400$

The values of the electrode currents are given in Table 1. The loci of potential maxima and minima, and the locations of the root points, are

shown in Fig. 4, whereas Fig. 5 is a corresponding drawing in which a physical picture of the behaviour of the equipotentials is traced out for the purpose of illustration.

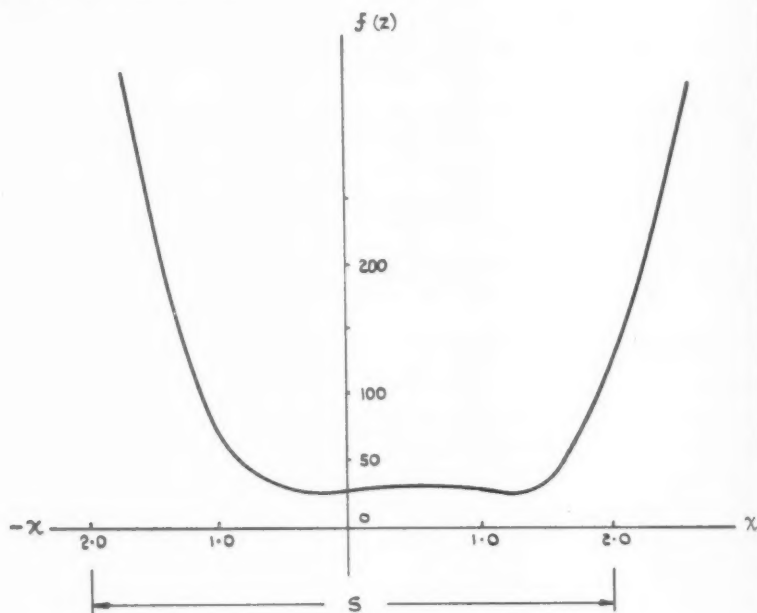


FIG. 3. The graph of

$$f(z) = z^8 - 3.012z^7 + 3.225z^6 + 1.0212z^5 + 6.9862z^4 - 21.889z^3 + 8.110z^2 + 5.9012z + 23.889$$

TABLE 1

Electrode	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
Position	2.0	1.5	1.0	0.5	0	-0.5	-1.0	-1.5	-2.0
Current (mA) $\left\{ \begin{array}{l} \text{source} \\ \text{sink} \end{array} \right.$	1	2.18	6.2	13.05	14.6	12.8	15.0	17.4	10.0

A comparison between the measured results and those given in (8) is shown in Table 2.

TABLE 2

Values given in ref. (8)	1.515 ± 10.618	1.51 ± 11.55	-0.48 ± 10.804	-1.043 ± 11.071
Measured values	1.512 ± 10.610	1.504 ± 11.566	-0.483 ± 10.79	-1.047 ± 11.067

The time required for locating the roots by the tank method is approximately 3 hours and the overall accuracy is about 1 per cent.

It should be pointed out that in Fig. 4 the loci of potential maxima and minima do not agree accurately with the configuration of the equipotentials

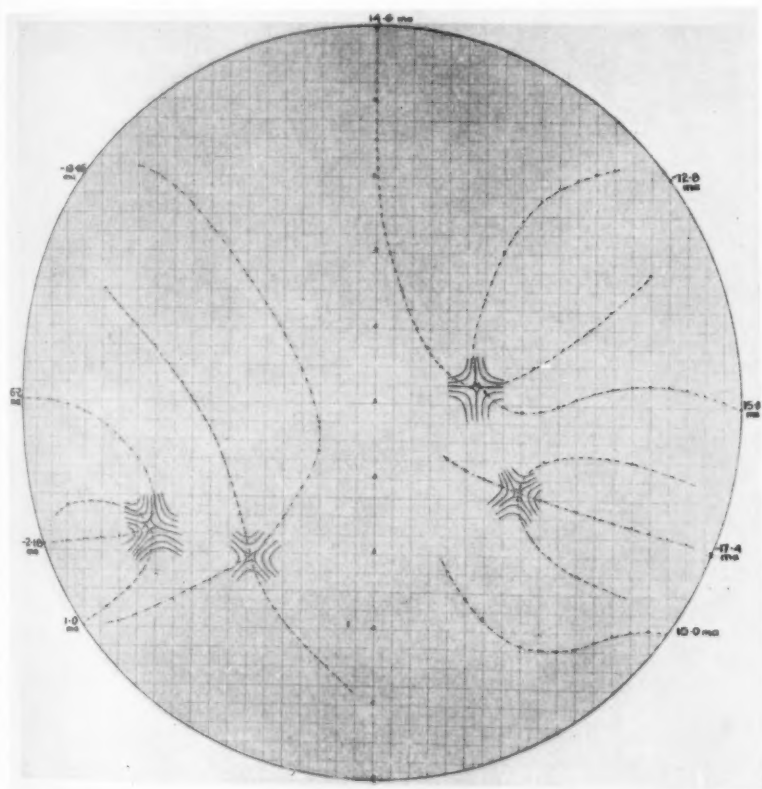


Fig. 4. The root locations of Example 1 on a single-layer electrolytic tank

in the vicinity of a root point. The explanation is that the tank surface was first searched by a rapid but approximate method by the searching probe. The probe was guided along the orthogonal lines on the graph paper at the bottom of the tank surface usually at 0.5 cm intervals. The maximum and minimum potentials, which are in fact the points of tangency to potential lines as shown in Fig. 5, were observed and their positions recorded. When a gradual curvature occurred along the path of measurement a longer mark was recorded than for a sharp curvature. The loci of potential

maxima and minima are lines constructed by joining these long or short marks, with correspondingly less or greater accuracy.

It cannot be expected, therefore, that these loci coincide very accurately with the actual configuration of the equipotentials. Also it must be noted

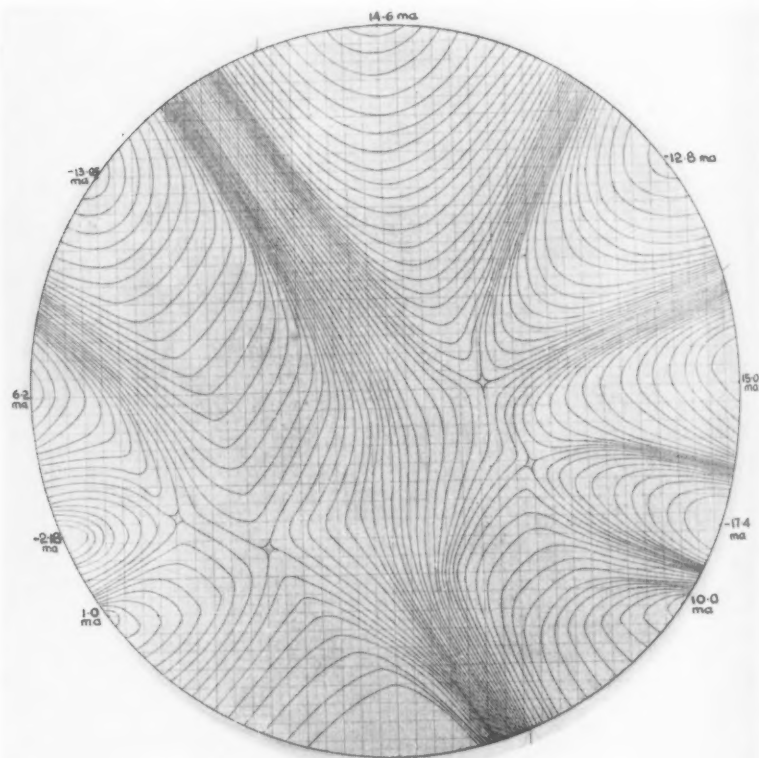
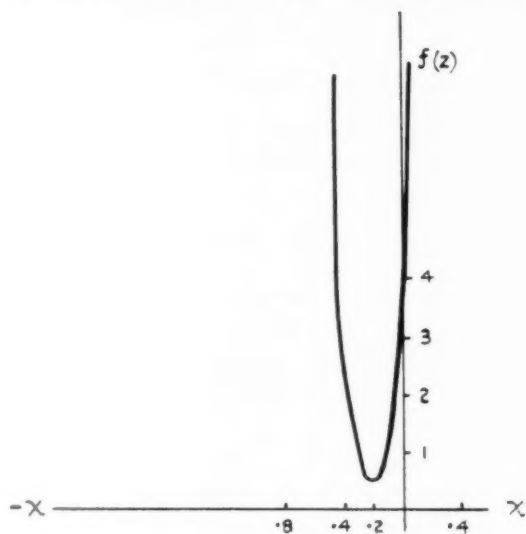


FIG. 5. Potential field representation of Example 1 corresponding to Fig. 4; lines are equipotentials

that the figure is produced directly from the experimental chart without modifications. Although these loci merely gave the clue to the region in which the final potential measurements should be made, the intersections of these loci do, nevertheless, give the root points accurately. In the next example attempts were made to locate the root points directly by the intersections of the loci without final potential measurements being made.

(a)



(b)

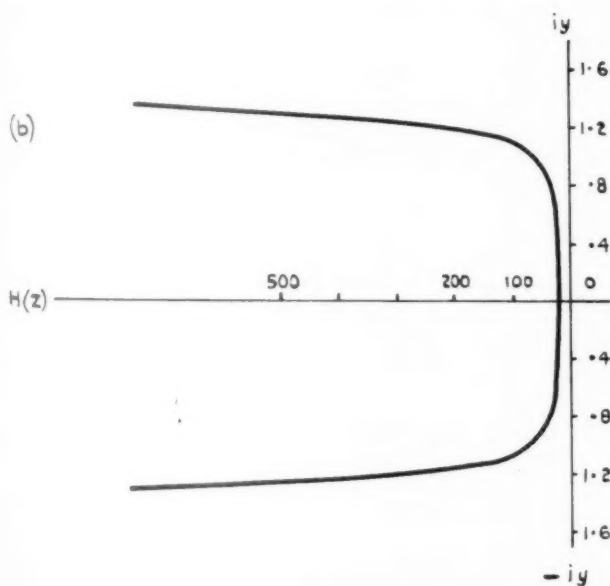


FIG. 6. $f(z)$ is given by equation (9), $H(z) = f(z)f(-z)$

Example 2

Here

$$f(z) = z^{16} + 2.40z^{15} + 9.639z^{14} + 16.922z^{13} + 34.673z^{12} + 45.465z^{11} + \\ + 60.264z^{10} + 59.018z^9 + 53.973z^8 + 38.643z^7 + 24.377z^6 + \\ + 12.149z^5 + 5.017z^4 + 1.573z^3 + 0.368z^2 + 0.056z + 0.00427. \quad (9)$$

Fig. 6 shows the graphs of $f(z)$ and $H(z) = f(z)f(-z)$ plotted against the x - and y -axes respectively. From this it is seen that the roots will occur within the region $x = 0, -0.5$, and $iy = 0, 1.6$, that is, in the neighbourhood along the y -axis. In order to set up current flow by electrodes along the imaginary axis, $H(z)$ is dealt with instead of $f(z)$; in this case, $2n+1$ electrodes are to be used, including one at $z = 0$. The values of the corresponding currents can be obtained after determining the partial fractions:

$$\frac{H(z)}{(z-iy_1)(z-iy_2)\dots(z-iy_{2n+1})} = \frac{k_1}{(z-iy_1)} + \frac{k_2}{(z-iy_2)} + \dots \quad (10)$$

Now, because the $2n$ roots of $H(z)$ are symmetrical about both the x - and y -axes, only one quadrant of the complex z -plane need be used for locating the roots, and the number of electrodes can also be reduced to $n+1$; the value of the current at $z = 0$ is one-half of that originally obtained from (10).

Electrode positions were chosen at $y = 0, \pm i(0.1, 0.2, 0.3, \dots, 1.5, 1.6)$. Only the loci of potential maxima and minima are shown in Fig. 7, and the positions of the roots were determined by the intersecting points of the two loci. The theoretical and measured values of the roots are compared in Table 3.

TABLE 3

Theoretical	$-0.210 \pm i0.104$	$-0.185 \pm i0.315$	$-0.174 \pm i0.490$	$-0.162 \pm i0.674$
Measured	$-0.191 \pm i0.094$	$-0.175 \pm i0.310$	$-0.149 \pm i0.478$	$-0.148 \pm i0.673$
Theoretical	$-0.148 \pm i0.917$	$-0.120 \pm i1.14$	$-0.105 \pm i1.31$	$-0.098 \pm i1.51$
Measured	$-0.111 \pm i0.917$	$-0.116 \pm i1.11$	$-0.095 \pm i1.32$	$-0.078 \pm i1.50$

The time required to locate all the eight conjugate pairs of roots was 2 hours with an accuracy of about 5 per cent.

4. Accuracy and advantages

The accuracy of solutions obtained by this method depends to some extent on the experience and judgement of the operator as well as on the intrinsic accuracy of the apparatus. However, the overall accuracy of the values of the roots obtained throughout the investigation was about 4 per cent. This

is considered to be a satisfactory result, since the root values obtained may be regarded as first approximations which may then be improved to any desired degree of accuracy by familiar numerical methods.

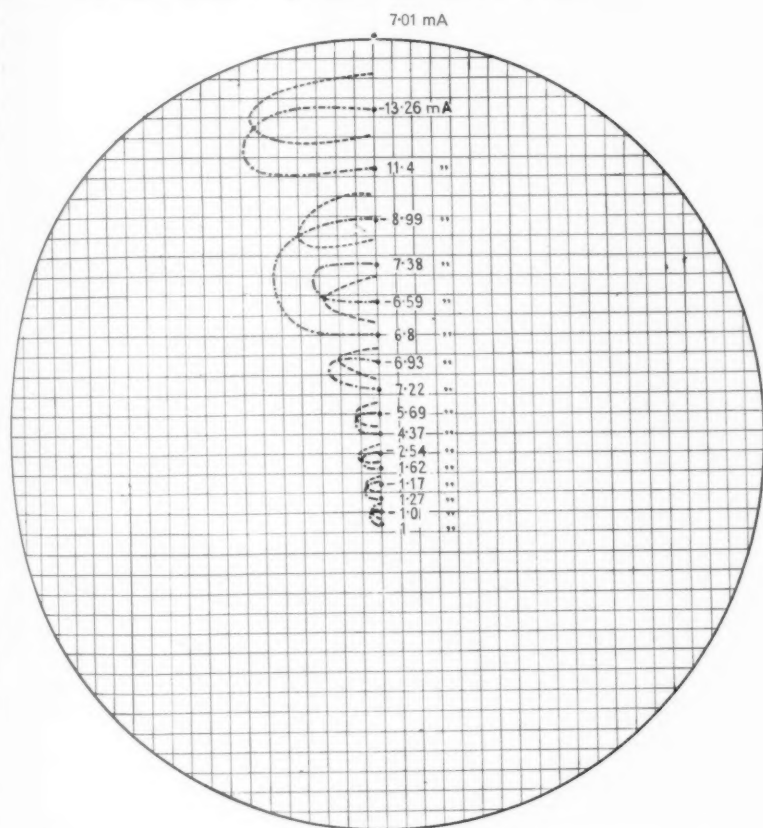


Fig. 7. Root locations of Example 2 on a single-layer electrolytic tank

Three main advantages result from the method described here, particularly when polynomials of high degree are involved:

(1) Provided the electrode arrangement is suitably selected, roots of polynomials of any degree can, in principle, be located accurately with approximately the same speed.

(2) All roots can be clearly visualized at once; this avoids the possibilities of root divergencies which are common if a solution is performed by ordinary numerical methods.

(3) When electric network problems are dealt with by means of the tank (9), pole and zero positions can be found directly on the tank surface by the present method.

When a polynomial is dealt with by means of the tank, the major part of the time is spent in the preliminary evaluation of the function and the subsequent determination of the residue values for the electrode arrangement by substitution in the relation given by equation (3). In the actual measurement, approximately half an hour was required to locate one root on the tank when equipotentials in the vicinity of the roots were traced out. The tracing of equipotentials may not, however, be necessary when only a first approximation is required. For then the locations of the roots can be determined with sufficient accuracy from the intersections of the loci of potential maxima and minima, as in Example 2. The nearness of the root points obtained in this way to those obtained by the tracing of the equipotentials is evident from Fig. 4. It may be remarked that the time required to find all the roots from the intersections of the loci of potential maxima and minima is no more than 1 hour.

The accuracy of the results described is limited by margins of experimental error imposed by the equipment used. This error arises in determining, usually at the intersection of the asymptotes of equipotentials, the exact position of a saddle point.

Four possible sources of such error are as follows:

(1) The degree of accuracy in the measurement of the current values required to set up the potential distribution was dependent upon the sensitivity and calibration of the milliammeter, variable resistances, power supply, etc. Although great care was taken, fluctuations of current values due to polarization at electrodes were inevitable in a long experiment.

(2) The degree of accuracy of measurements was dependent upon the sensitivity and calibration of the valve-voltmeter.

(3) In the actual determination of the equipotentials, the use of 0.5 mm diameter platinum wire as measuring probe limited the number of potential measurements (for tracing equipotentials) that could be made within a unit area around the zero, and consequently the certainty of definition of the null region as provided by the tracing of the equipotentials.

(4) The thickness of the probe may also have affected the degree of accuracy with which, within the null region as defined by the tracing of the equipotentials, the exact position of the root point was determined.

In addition, the equations dealt with in the foregoing experiments include many of a nature that gave rise to unfavourable conditions for solution by the method described. Such unfavourable conditions include, for example,

when the roots are (a) very far from both the x - and y -axes; (b) cluster together;† (c) vastly different in moduli.

In these experiments, despite the use for each equation of a number of electrode positionings in order to find for each the optimum electrode positions, the residues obtained exhibited vast (highly unfavourable) ratios.

In the next section it will be shown that the greater the ratios exhibited by the residues, the more unfavourable the conditions for the method described.

5. Justification of the choice of arbitrary roots x_1, x_2 , etc., indicated by the values of residues

The determination of the values of the residues from the partial fractions obtained by expanding $G(z)$, as in equation (3), gives the values for the points x_s giving rise to the potential distribution corresponding to the polynomial $f(z)$ of (1). The process is straightforward yet sometimes laborious, particularly for high degree polynomials. As the roots in equation (2) are chosen arbitrarily, many different electrode arrangements can be obtained, but only that close to the ideal condition will be of interest. The ideal condition is that the electrodes should be close to the root points: all the electrodes should be placed within the region S chosen, and the ratio of the distance between the furthest root and the axis where the electrodes are positioned and the distance between electrodes should preferably be not more than 4 to 1. Because the ratio does not exceed 4:1, the ratio of the highest to the lowest value of the residues will generally be of the order of not more than 25 to 1. It is found that the lower the ratio between the values of the residues, the more favourable the conditions for obtaining solutions. When a set of residues is obtained, one can then tell whether such an arrangement of electrodes is suitable. If not, fresh electrode positions are chosen. In this way, electrode arrangements can be justified by observing the values of the residues until a satisfactory set is obtained. This eliminates a lot of unnecessary trial on the electrolytic tank.

When a set of residues is obtained having values of vast ratio, two possibilities may exist. It may be either that electrode positions were not suitably chosen or that the roots are unfavourably distributed. There is no rule to improve the positioning of the electrodes, but experience of judging the residue values is a great advantage and this may help to determine whether such a polynomial can be solved by the tank method at all without the tank needing to be set up.

† When the roots are clustered together it is impossible in practice to differentiate one zero from another because the current intensity vanishes in their vicinity and only the region where they occur can be located.

6. Conclusion

The application of this method for factorizing polynomials of higher order than the 16th, which is the highest attempted in this work, is not impossible; and theoretically the method can be extended to solve equations of any degree. It is hoped that the electrolytic tank will prove extremely valuable in the analysis and synthesis of electric networks and theoretical analysis of those physical and engineering problems which may be represented as polynomials.

7. Acknowledgements

The author wishes to thank Professor Willis Jackson for his permission to allow this research to be undertaken in the Electrical Engineering Department, Imperial College; also Dr. E. C. Cherry, who suggested the subject and subsequently supervised the work, and Dr. A. R. Boothroyd whose assistance throughout was invaluable.

REFERENCES

1. FELIX LUCAS, 'Détermination électrique des lignes isodynamiques d'un polynôme quelconque', *C. R. Acad. Sci., Paris*, **106** (1888) 587-9.
'Résolution immédiate des équations au moyen de l'électricité', *ibid.* 645-8.
2. CHARLES H. ZIMMERMAN, *An Analysis of Lateral Stability in Power-off Flight with Charts for use in Design* (N.A.C.A. Report 589, 1937).
3. W. V. LYON, 'Note on a method of evaluating the complex roots of a quartic equation', *J. Math. Phys.* **3** (1924) 188-90.
4. Y. H. KU, 'Note on a method of evaluating the complex roots of a quartic equation', *ibid.* **5** (1926) 126-8.
5. L. F. WOODRUFF, 'Note on a method of evaluating the complex roots of sixth and higher order equations', *ibid.* **4** (1925) 164-6.
6. R. E. DOHARTY and E. G. KELLER, *Mathematics of Modern Engineering* (John Wiley and Sons, N.Y., 1936).
7. F. L. HITCHCOCK, 'Finding complex roots of algebraic equations', *J. Math. Phys.* **17** (1938) 55-58.
8. S. N. LIN, 'A method of successive approximation of evaluating the real and complex roots of cubic and higher-order equations', *ibid.* **20** (1941) 231-42.
9. A. R. BOOTHROYD, E. C. CHERRY, and R. MAKAR, 'An electrolytic tank for the measurement of steady-state response, transient response and allied properties of networks', *Proc. I.E.E.* **96** (1949) 163.
10. S. L. BROWN and L. L. WHEELER, 'A mechanical method for graphical solution of polynomials', *Franklin Inst. J.* **231** (1941) 223-43.
11. R. L. DIETZOLD, 'Isograph—A mechanical root-finder', *Bell Laboratory*, **16** (1937) 130-4.
12. BYRON O. MARSHALL, JR., 'The electronic isograph for roots of polynomials', *J. App. Phys.* **21** (1950) 307-12.

MACHINE

her order
possible;
ns of any
valuable
analysis
sented as

rmission
ineering
sted the
othroyd

un poly-

645-8,
ff *Flight*

a quartic

quartic

of sixth

g (John

r. *Math.*

real and

31-42.

for the

properties

solution

ory, 16

omials',

THE QUARTERLY JOURNAL OF MECHANICS AND APPLIED MATHEMATICS

VOLUME X

PART 3

AUGUST 1957

CONTENTS

A. CEMAL ERINGEN: Elasto-dynamic Problem concerning the Spherical Cavity	257
DAPHNE G. PADFIELD and JEAN SIDA: The Indentation of a Thick Sheet of Elastic Material by a Rigid Cylinder	271
A. M. BINNIE: Experiments on the Slow Swirling Flow of a Viscous Liquid through a Tube	276
H. J. DAVIES and A. J. ROSS: A Jet deflected from the Lower Surface of an Aerofoil	291
O. E. POTTER: Laminar Boundary Layers at the Interface of Co-current Parallel Streams	302
J. C. COOKE: The Flow of Fluid along Cylinders	312
E. WILD: Electromagnetic Waves in Nearly Periodic Structures	322
A. W. GILLIES: On a Class of Differential Equation governing Non-linear Vibrations	342
A. W. BABISTER: Response Functions of Linear Systems with Constant Coefficients having One Degree of Freedom	360
S. K. IP: An Electrolytic Tank as an Analogue Computing Machine for factorizing High Degree Polynomials	369

The Editorial Board gratefully acknowledge the support given by: Blackburn & General Aircraft Limited; Bristol Aeroplane Company; Courtaulds Scientific and Educational Trust Fund; English Electric Company; Hawker Siddeley Group Limited; Vickers-Armstrongs (Aircraft) Limited.

The publishers are signatories to the Fair Copying Declaration in respect of this journal. Details of the Declaration may be obtained from the offices of the Royal Society upon application.

ICS

1957

257

271

276

291

302

312

322

342

360

369

Gen-
onal
ers-